

ソーシャルキュレーションデータからの観点抽出と画像検索への応用

金 応教[†] 山本 岳洋[†] 田中 克己[†]

[†] 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 本研究では、画像に対する観点を、ユーザが画像を見てどう捉えるかを表す概念として定義する。近年、多くのユーザに利用されている Pinterest に代表される、画像のソーシャルキュレーションサービスでは、様々な人々がある観点で画像を集め、整理し、公開することができる。このようにして構築されたキュレーションデータには、画像に対するユーザの観点が強く表れていると考えられる。本研究では、多くの人々が類似した捉え方を持つ画像を探せるという観点に基づく画像検索において、ソーシャルキュレーションデータが有用であると考え、そのデータから観点を抽出し、画像の観点をクエリとするランキング手法について提案し、“cute”や“beautiful”といった、画像から受ける印象に関連する観点を表すクエリに対して、提案手法の精度を検証する。

キーワード 画像検索, ソーシャルキュレーション

1. はじめに

ウェブの発展に伴い、ウェブ上に存在するコンテンツの数も増加し、人々はそのコンテンツに接する機会も増加してきた。例えば、画像の場合は、スマートフォンやタブレットなどが人々の身近に定着することにより、デバイスに付属されている携帯カメラを使って簡単に写真を撮って、ウェブ上の画像共有サービスを通じてそれらの画像を容易にアップロードし、共有することができる。画像共有サービスとしては、Flickr^(注1)や Instagram^(注2)などが流行しており、人々はこれらの画像共有サービスを用いて膨大な数の画像にアクセスすることができる。

画像共有サービスでは、画像をアップロードする際に、ユーザが画像に何かの付加情報を付与することができる。例えば、Flickr の場合、ユーザがアップロードする写真を説明するために、その画像のタイトルを決め、説明文を作成し、簡潔な単語で表されるタグを付与することができる。ユーザは、画像に対する付加情報に基づいた検索システムを使い、自分の検索意図を表すクエリを入力して必要な画像を見つけることができる。

ユーザが画像検索を行うとき、ユーザは欲しいと思う画像に対して特有の捉え方を持っており、その捉え方に基づいて検索目的に合致していると考えられるクエリを選択する。本稿では、画像に対する人の捉え方を、画像の観点と定義する。つまり、観点というものは、人が画像を見てどう感じ取るかを表すものである。例えば、ユーザが“鳥居が写っている”、“祇園祭が写っている”という、画像に直接写っているオブジェクトやイベントに関する観点を持っている場合、画像検索のためのクエリとして、“Torii”、“Gion Matsuri”を選択する。また、“かわいいと感じられる”というユーザ個人の感想や捉え方に関連する観点の場合は、クエリとして“cute”という個人の捉え方を表すキーワードを用いる。

従来の画像検索システムは、前述した説明文やタグなどの画像に対する付加情報が、画像の観点を表していると思われ、それらを検索システムの基盤となるデータとして用いている。また、一部の画像検索システムでは、画像の RGB ヒストグラム、HSV ヒストグラム [1]、Bag of Visual Words [2] などの画像特徴量を用い、ユーザが選択した画像の特徴量に基づいて類似している画像を探すこともできる。

しかし、これらのデータは、画像が先にあるとその画像に対して付与されたものなので、そのデータが画像を見る人々の捉え方を表す観点を反映しているとは言い難い。例えば、“Kyoto”という語がタグとして付加されている画像は、“京都の風景が写っている画像”である場合がほとんどで、はっきり京都が写っているわけではないが“京都が連想できる画像”には、“Kyoto”というタグが付加されていない場合が多い。

また、多くの人々が同じ観点を捉える画像に対しても、前述のデータだけでは、その観点に対する度合いを求めることが困難である。例えば、“京都といえばこれ、という画像”、“誰が見ても京都と捉える画像”などの場合は、単に京都が写っている画像と同じく、“Kyoto”というタグが付いているだけの場合が多いので、そのデータを用いて“京都らしさ”の度合いを計算することができない。

特に、抽象的、主観的、感覚的な観点であるほど、その観点は画像を見る人々の考えに大きく依存することになるので、前述した画像の付加情報だけでその観点を表現するのは難しい。従って、従来の画像検索システムで、画像から受ける印象に関連する観点を捉えて、“かわいいと思える画像”、“美しいと感じる画像”などを探るとき、“cute”、“beautiful”などのクエリで画像検索を行っても、本当にそれらの観点を捉えられる画像が得られない場合が多い。つまり、上記のような抽象的、主観的、感覚的なクエリでは、あまり良い画像検索結果は得られない。

そこで、本研究ではキュレーションに注目する。情報分野におけるキュレーションとは、情報コンテンツを集めて、整理し、人に見せることを意味し、そのキュレーションを行う人のこと

(注1): <https://www.flickr.com/>

(注2): <http://instagram.com/>

をキュレーターと呼ぶ。キュレーションを行う際、集められたコンテンツにはキュレーターの意図という新たな価値が生まれ、さらに、要約、解説、評価などの新たな情報が、キュレーターによってキュレーションデータとして付加される。従って、キュレーションデータは、従来のコンテンツが先にあるの付加情報とは異なる、キュレーターが先にあるの情報として扱うこともできる。

特に、ソーシャルキュレーションとは、様々な人々が参加するキュレーションのことであり、“一緒にキュレーションを行う”、“誰でもキュレーターになれる”といった、2つの意味で考えることができる。どちらも様々な人々が参加するので、キュレーションデータには様々なキュレーターによる観点が含まれることになる。本研究では、ソーシャルキュレーションを後者の意味として捉え、一般の人々の観点がそれぞれのキュレーションデータに含まれ、その様々なキュレーションデータを扱うことで、多様な人々の多様な観点をデータとして扱うことができると考える。

画像コンテンツのソーシャルキュレーションサービスとしては、最近 Pinterest^(注3) というサービスが流行している。Pinterest では、ユーザが Web 上の画像に対するビジュアルブックマークと、それらのビジュアルブックマークを集めたコレクションを自由に作成し、整理、共有することができる。本稿では、ユーザによってつけられるコレクション名に注目し、このコレクション名に画像に対するユーザの観点が含まれていると考え、ある画像に対する様々なコレクション名から様々なユーザの観点を抽出し、多くの人々が類似した捉え方を持つ画像を探るような、人々の観点到合う画像検索のための応用として、観点のデータに基づいて画像検索結果をランキングする手法を提案する。

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 Pinterest の構造を説明し、その構造に基づいたグラフの求め方を紹介する。4章では印象を表すクエリによって得られる画像検索結果を、ソーシャルキュレーションデータに基づいてランキングする手法を提案する。5章では提案手法に対する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 抽象的なクエリでの画像検索

抽象的なクエリを用いた画像検索を改善に焦点を当てている研究としては、Kato らの研究がある [3]。Kato らは、ソーシャル画像共有サービスのタグ情報を用いて、抽象的な語からの具体的な語への対応づけを行い、その対応づけを用いて抽象的なクエリでの画像検索の改善を試みている。しかし、この研究では大きな改善は見られなかったため、更なる改善の余地を示唆しており、本研究では、抽象的なクエリに対しては、タグという画像主体の情報源より、キュレーション情報というユーザ主

体の情報源の方が効果的であると考え、特に印象を表すクエリに注目し、キュレーション情報を用いることで印象を表すクエリでの画像検索を改善できるのではと考えている。

2.2 Pinterest を対象とした研究

Pinterest を分析対象とした研究としては、Gilbert らの研究や Linder らの研究、Kimura らの研究がある。Gilbert らの研究では Pinterest のユーザとデータの統計的な分析を行っており [5]、Linder らの研究ではユーザとのインタビューを通じて、アイデア作りのツールとしての Pinterest を語っている [6]。Kimura らの研究では Pinterest のデータから画像の文脈を抽出し、それを新たな画像特徴量として利用できる可能性を示唆している [4]。これらの研究では、Pinterest という情報源の分析と利用可能性に焦点を置いているが、本研究では、Pinterest を画像検索でどう応用できるかに焦点をおくという点で、彼らの研究とは大きく異なる。本研究で提案するキュレーション情報を用いた画像検索結果のランキング手法を利用することで、よりユーザの観点の沿った画像検索結果を提示できるのではと考えている。

3. Pinterest

3.1 Pinterest とは

Pinterest とは、ビジュアルコンテンツを対象としたソーシャルブックマークサービスであり、一般のユーザがビジュアルコンテンツを収集するという意味ではソーシャルキュレーションサービスでもある。Pinterest では、ユーザが Web 上のビジュアルコンテンツに対するビジュアルブックマークを作成でき、このビジュアルブックマークのことをピン (Pin)、ブックマークを作成する行為のことをピンと呼ぶ。また、Pinterest では、これらのピンをコレクションとして集めることができ、このピンのコレクションのことをボード (Board) と呼ぶ。ピンの説明文は、ピンする際に元の画像の説明文が自動で入力されているが、ピンするユーザはこのピンの説明文を自由に修正することができる。ボードの名前、ボードの説明文は、元の画像には存在しないものなので、ユーザが自由に決めることになる。

Web 上の画像をピンする際、基本的に、ユーザは決まったボードにそのピンを入れることになる。ピンするときに新しくボードを作成することもできるが、一つのピンのためにわざわざボードを作成するというより、予めユーザが決めていた範疇で画像を探してピンすることになるので、この場合もピンするときに既にボードが決まっていると言っても間違いではない。このような側面で Pinterest を見ると、ピンの情報より、ボードの情報の方がユーザの意図が入りやすくなっており、ユーザの観点が含まれているデータとして適していると考えられる。

Pinterest のピンにはもう一つ特徴があり、それがリピン (Repin) というものの存在である。リピンとは、Pinterest 上のピンに対して作成したピンのことである。つまり、Pinterest は、ピンをピンするとリピンというピンが作成される、再帰的な構造になっている。このリピンというものは、元のピンに影響されて作成したものであるとも考えられるので、元のピンとリピンの間の重要度の伝播などとしても利用できる。他にも、

(注3): <https://www.pinterest.com/>



図 1 ピンの例

Pinterest のユーザは、ピンに対していいね (Like) という評価をすることも、ボードやユーザをフォロー (Follow) することもできるが、本研究ではそれらの機能は割愛する。

図 1 は Pinterest 上のあるピンのページ^(注4)である。ピンのページの中央にはピン元の画像が表示され、右側にはこのピンが含まれているボードの名前とそのボードの作成者、そのボードに含まれているいくつかのピンの画像が表示されている。ピンの画像のすぐ下には元の画像がどのドメインから得られたものであるかが分かり、その部分はリンク処理が施されているので、クリックして元画像のページを開くことができる。もう少し下に下がると、ピンに対するコメントが表示されており、ピンするときに入力したピンの説明文は、ピンしたユーザによる最初のコメントとして登録される。コメントの下にはこのピンを作成したユーザが Added by として表示され、このピンがリピンであると、Added by の右に Via として元のピンの作成者が表示され、このピンが Web 上の画像を直接ピンしたものであると、Added by の左にこのピンが含まれるボードが Added to として表示される。画像の上には、様々な機能が備わっており、このピンをリピンしたい場合は左の Pin it ボタンを使って可能で、ボタンの隣の数字をクリックすると、このピンに対するリピンが含まれるボードが見られる。

図 2 は Pinterest 上のあるボードのページ^(注5)である。まず検索バーを除いた最上部にはボード名が表示され、もしボードの説明文が存在する場合はボード名のすぐ下に表示される。ボード名とボード説明文の下にはボードに参加しているユーザ (Pinterest では、複数のユーザが一つのボードにピンを追加することができる)、ボードに含まれるピンの数、ボードをフォローしているフォロワー (Follower) の数が表示される。それらのボード情報の下には、実際にこのボードの中に含まれるピン



図 2 ボードの例

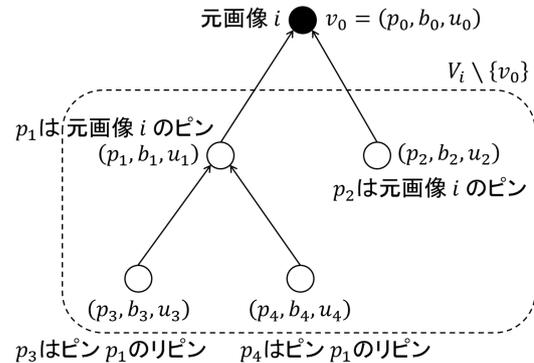


図 3 ピングラフの例

が、ピンの元画像、ピンの説明文、ピン元のドメイン情報という形でともに表示される。もしこのボードに複数のユーザが参加している場合は、ピンしたユーザの情報もともに表示される。

3.2 ピングラフ

本章で述べた Pinterest のピンの再帰的な構造は、グラフとして表すことができる。ある一つの画像 i に対して、その画像 i を元画像とする全てのピンを集めたピンの集合 P_i を求め、各ピン $p \in P_i$ のキュレーションデータとして、ピン p が含まれているボードのボード名 b 、ピン p をボードに追加したユーザ u のデータを取得すると、画像 i を元画像とするピンの構造とキュレーションデータを表すグラフを作成できる。

一つの画像 i を元画像とするピンの集合 P_i に対し、各ピン $p \in P_i$ とそのボード名 b 、ユーザ u の組をノードとし、“ピンする”というピンとピン元の間を、ピンのノードからピン元のノードへの有向エッジとすると、ただ一つのピングラフが得られる。

上記のプロセスで得られるある画像に対するピングラフは、全てのノードが連結されているのではなく、いくつかの小さな部分グラフで構成されているので、きれいな形のグラフとは言えない。そこで、仮想のピン p_0 、仮想のボード名 b_0 、仮想のユーザ u_0 を考え、元画像を形式的にノード (p_0, b_0, u_0) としてピングラフに追加し、元画像の直接のピンのノードからピン元である元画像のノードへの有向エッジを追加すると、最終的に全てのノードが連結されたただ一つの大きなグラフが得られることになる。この追加によって元画像を表す特殊なノードが一つ存在することになるが、エッジの意味はピンとピン元の間を関連を表すというものとして変化していない。

(注4): <http://www.pinterest.com/pin/291397038362516563/>

(注5): <http://www.pinterest.com/pen0421/adelie-penguin/>

以上により、一つの画像に対して図3のようなピングラフが作成できる。画像 i に対するピンの集合 P_i に対し、元画像の仮想ピン p_0 を追加してピン集合 $P'_i = P_i \cup \{p_0\}$ に拡張すると、画像 i に対するピングラフ G_i は以下の式で定義される。

$$G_i = (V_i, E_i)$$

$$V_i = \{(p, b, u) | p \in P'_i, b \text{ is a board name, } u \text{ is a user}\}$$

$$E_i = \{(v_j, v_k) | v_j, v_k \in V_i, p_j \text{ is a pin of } p_k\}$$

ここで、以降で用いるため、元画像を除いたピンだけのノード集合を、 $\tilde{V}_i = \{(p, b, u) | p \in P_i, b \text{ is a board name, } u \text{ is a user}\} = V_i \setminus \{v_0\}$ と定義しておくことにする。

得られた大きなピングラフには、各ノードにピンに対するキュレーションデータが含まれているので、このピングラフを観点の塊としてみなすことができ、このピングラフを画像検索に活用することで、様々なユーザの観点を画像検索に反映でき、人々の感覚に合う画像検索システムを実現できると考えられる。

4. ソーシャルキュレーションデータに基づく画像検索結果のランキング

本章では、Pinterest のソーシャルキュレーションデータに基づいて、クエリに対する画像検索結果をランキングする手法について記述する。

まず、本手法の問題設定を述べる。入力としては、画像集合 I 、検索クエリ q と、前章で述べたピングラフの集合 $G_I = \{G_i | i \in I\}$ が与えられ、出力として、スコア付き画像集合 $S(I, q) = \{(i, s(i, q)) | i \in I, s(i, q) \text{ is a score}\}$ を得るのが、本手法に対する問題設定になる。つまり、ソーシャルキュレーションデータを用いて、画像集合とクエリに対するランキング結果を得ることが、本手法の目的である。

本手法では、まず、ピングラフのボード名のデータから観点を抽出する。次に、抽出した観点とピングラフのピンの構造に基づいて、クエリが表す観点に対する各画像の観点の適合性をスコアとして求め、そのスコアに基づいて画像集合をランキングする。

4.1 観点抽出

まず行うべきことは、ボード名から観点を抽出することである。本手法では、ピングラフの各ボード名に自然言語処理を行っていくつかの語に切り出し、語自体に意味が含まれていない文法のための語を取り除き、名詞、動詞、形容詞、副詞の語のみをボード名に含まれる観点として扱う。これにより、一つのボード名 b からユーザの観点を表す語のみで構成される語集合 T_b が得られることになる。以降は、各ノードが表すピンのボード名から抽出された観点を表す語集合を、そのノードの観点として扱う。

4.2 スコアリング

クエリに対して画像集合をランキングするためには、各画像に順位を付けるための計算を行う必要がある。本稿では、クエリに対する各画像の観点の適合性を、各画像のクエリに対するスコアとして計算し、そのスコアをランキングに用いる。

本稿では、一つ画像のクエリに対するスコアを計算するため、

その画像のピングラフに含まれる各ノードの観点に対し、以下を考慮する。

クエリとの適合性 クエリと類似した観点を持つノードが多いほど、元画像はクエリとの適合性が高い。観点はあるユーザの捉え方を表すので、多くのユーザが一つの画像に対して類似した観点を持っていると、その画像はその観点に大きく適合していると言える。

ピン元に対する独立性 ノードが表すピンの、ピン元に対する独立性が高いほど、そのノードの観点はクエリに対する重要度が高い。ここで、ピン元に対するピンの独立性というものは、ピンのボード名による観点が、ピン元の付加データの影響をどれほど受けていないのかを表す概念である。つまり、他のデータからの影響を受けていない観点は、高い価値を持っているということである。本稿では、このピン元に対する独立性として、2つの考え方を提案する。

(1) 全てのリピンは、独立性が低い。リピンには必ず元のピンが存在するので、リピンがその元のピンの影響を必ず受けていると考えることができる。

(2) リピンを表すノードの観点が、元のピンを表すノードの観点と類似していると、そのリピンは独立性が低い。これは、あるピンに対してリピンを作成するとき、リピンをどのボードに入れるかは元のピンのボード名の影響を受ける場合があり、両者のボード名が類似していると、強い影響を受けたと考えることができる。逆に、類似していない場合は、元のピンの影響をまったく受けなかったと考えることができる。

上記のアイデアから、スコアを計算するための、以下のようないくつかの手法を考えることができる。

(1) 適合性のみを考慮したスコア計算手法が考えられる。クエリと類似した観点を持つノードが多いほど、高いスコアになるようにすれば良い。ただし、単にノードの数が多いだけでスコアが高くなるのではなく、多様な観点が含まれるより類似した観点が多く含まれる方が適合性が高い、という考え方を考慮し、ノードの数による正規化を行う。クエリ q に対する適合性のみを考慮した画像 i のスコアの式を以下に示す。

$$s(i, q) = \frac{1}{|\tilde{V}_i|} \sum_{(p, b, u) \in \tilde{V}_i} \text{Sim}(q, b) \quad (1)$$

ここで、クエリ q とボード名 b の類似度 $\text{Sim}(q, b)$ は、クエリの語集合を T_q 、ボード名から抽出した観点の語集合を T_b とし、以下のように定義する。

$$\text{Sim}(q, b) = \max_{t_q \in T_q, t_b \in T_b} \text{Sim}(t_q, t_b)$$

語の間の類似度としては、WordNet Similarity の Path Similarity [7] を用いることにする。以降、クエリとボード名の類似度としては、上記の式を用い、ボード名の中の類似度に対しても、観点の語集合を用いて同様に上記の式を使用する。

(2) 適合性と独立性の (1) を考慮したスコア計算手法が考えられる。クエリとの適合性の考え方は、基本的に式 (1) と同じであるが、リピンのノードの場合は独立性が低いとみなし、適合性に与える影響を低くする重みを付加する。クエリ q に対

する適合性とピンの独立性の (1) を考慮した画像 i のスコアの式を以下に示す .

$$s(i, q) = \frac{1}{|\tilde{V}_i|} \sum_{(p,b,u) \in \tilde{V}_i} w_p \cdot \text{Sim}(q, b) \quad (2)$$

$$w_p = \begin{cases} 1 & (p \in P_i \setminus R_i) \\ \alpha & (p \in R_i) \end{cases}$$

ここで, R_i は画像 i を元画像とするリピンの集合で, α は $0 < \alpha < 1$ を満たす定数である .

(3) 適合性と独立性の (2) を考慮したスコア計算手法が考えられる . 基本的な考え方は, 式 (2) と同じであるが, リピンの独立性を, ピン元との類似度を用いて計算するので, 重みの計算方法が式 (2) とは異なる . クエリ q に対する適合性とピンの独立性の (2) を考慮した画像 i のスコアの式を以下に示す .

$$s(i, q) = \frac{1}{|\tilde{V}_i|} \sum_{(p,b,u) \in \tilde{V}_i} w_{p,b} \cdot \text{Sim}(q, b) \quad (3)$$

$$w_{p,b} = \begin{cases} 1 & (p \in P_i \setminus R_i) \\ 1 - (1 - \alpha) \text{Sim}(b, \hat{b}) & (p \in R_i) \end{cases}$$

ここで, R_i は画像 i を元画像とするリピンの集合, \hat{b} はピン p の元のピンのボード名で, α は $0 < \alpha < 1$ を満たす定数である .

以上の各手法で計算したスコアを用いると, 画像集合をクエリに対してランキングすることができる . スコアが高い画像であるほど, 上位にランキングされ, クエリが表す観点に適合していると考えることができる .

5. 評価

本稿では, ソーシャルキュレーションデータに基づいて, 各画像に対してクエリが表す観点との適合性を意味するスコアを計算し, そのスコアを用いて画像集合をランキングする手法を提案している .

そこで, ソーシャルキュレーションデータの有用性を評価し, 提案したランキング手法の有用性を検証するため, 本手法を Python を用いて実装し, 2 種類の実験を行った .

5.1 実験設定

今回の実験では, 画像集合の大きなドメインを Flickr と設定し, 扱う画像としてもその中で Pinterest にピンされている画像だけに限定した . まず, ソーシャルキュレーションデータを取得するため, 元画像のドメインを Flickr とする Pinterest 上のデータをクロールした . また, Flickr が提供している API^[注6]を用いて, 集めた Pinterest のピンの元画像に対するタグなどの付加データを取得した .

実験のクエリとしては, 従来の画像検索ではあまり良い精度が得られないとされている形容詞の中で, クロールした Pinterest のボード名の中で頻りに現れる, “cute”, “beautiful”, “adorable”, “amazing”, “funny”, “cool”, “awesome”, “furry”, “sweet”, “happy” という, 印象を表す 10 個の形容詞クエリを使用した . 各クエリに対して, そのクエリが 1 回で

表 1 ソーシャルキュレーションデータの有用性に対する評価結果

適合判定	only board	only tag
1 点以上を適合	0.71	0.65
2 点以上を適合	0.33	0.26

もボード名に含まれる画像を全て取得し, ピンの数が多い上位 50 件の画像に対しては各画像のピングラフを作成した .

ピングラフを作成した画像に対しては, 4 章の提案手法を適用した . その際, 自然言語処理のツールとしては, Python の nltk モジュール^[注7]と, Stanford POS tagger^[注8]を使用し, スコアの計算におけるパラメータである式 (2) と (3) の α としては, $\alpha = 0.1$ を用いた .

5.2 ソーシャルキュレーションデータの有用性の評価

まず, 印象を表すクエリに対して, 本当にソーシャルキュレーションデータが従来の画像の付加データより優れているのかを評価した .

比較評価のために, 集めた全画像集合の中から, 設定した各クエリに対して以下の条件を満たす画像をランダムに 10 件ずつ取得した .

(1) only board: ボード名だけにクエリが含まれる

(2) only tag: タグだけにクエリが含まれる

各クエリに対して, 計 20 件の取得された画像をシャッフルし, 画像とクエリを提示しながら, 各画像を見てクエリが表す概念としてどれほどそうだと感じるかを評価者に質問し, “強く感じる” を 2 点, “そこそこ感じる” を 1 点, “感じない” を 0 点とする, 3 値で評価してもらった .

評価者の回答から, 1 点以上を適合とする, 2 点以上を適合とするという 2 つの場合を考え, 各クエリと各条件に対して画像集合の正解の割合を表す適合率を求め, 全クエリに対する適合率の平均を求めた . その評価で得られた結果を表 1 に示す .

表 1 の評価結果を見ると, 1 点以上を適合とする場合と, 2 点以上を適合とする場合の両方で, ボード名にクエリが含まれている場合の方が, タグにクエリが含まれている場合より高い精度が得られた . これは, 印象を表すクエリに対しては, タグよりボード名の方が有効であることを示している . また, 2 点以上を適合とする場合は精度の差がより大きくなることは, タグの場合は, 観点に対して適合しているかが曖昧な場合でも, その観点がタグとして付けられることが多いことに対し, ボードの場合は, 本当にボード名が表す観点に対して適合しているものだけを, そのボードのピンとして入れる傾向があるためであると思われる .

しかし, ボード名の場合とタグの場合の精度の差はあまり大きくなかった . ボード名に画像を見る人の観点が含まれやすいとは言っても, その観点というものは人によって様々で, ある人が特定の観点を持っているとしても, 多くの人はその観点を正しくないと思う場合も存在する . 本節の評価では, クエリが含まれているボード名が一つでも存在すれば評価対象の画像と

[注7]: <http://www.nltk.org/>

[注8]: <http://nlp.stanford.edu/software/tagger.shtml>

[注6]: <https://www.flickr.com/services/api/>

なりうるので、評価する画像集合の中には、前述した大多数が同じ観点を持つ画像はあまり存在しなく、評価結果であるボード名の精度の値も低くなってしまふ。従って、人々の観点に適した画像を検索するユーザに提示するためには、ボード名をそのまま用いるのではなく、ボード名を用いた適切な手法が必要である。

5.3 ランキング手法の有用性の評価

次は、印象を表すクエリに対して、提案したランキング手法が本当に有効であるか、従来の付加データであるタグを用いた手法よりどれほど優れているかを評価した。

比較評価のために、まず従来の付加データであるタグを用いるベースライン手法を設定した。Flickr の API で集めた画像集合の中から、各クエリに対してそのクエリが 1 回でもタグに含まれる画像を、タグが多い順で並べた上位 50 件の画像を取得し、式 (1) をタグに合わせて修正した以下の式 (4) でスコアを計算し、スコアが高い順にランキングした。

$$s(i, q) = \frac{1}{|\mathcal{T}_i|} \sum_{\tau \in \mathcal{T}_i} \text{Sim}(q, \tau) \quad (4)$$

ここで、 \mathcal{T}_i は画像 i に対するタグの集合である。

ベースラインと比較する提案手法としては、4 章で述べた、スコアを計算し、スコアが高い順にランキングする 3 つの手法を採択した。この 3 つのランキング手法を適用する画像集合としては、各クエリに対して既にピングラフを作成した 50 件の画像を用いた。

クエリが表す観点に対して画像をランキングする、以上の比較手法をまとめたものを以下に表す。

- (1) ベースライン：タグに基づいて式 (4) でランキング
- (2) 適合性：ボード名に基づいて式 (1) でランキング
- (3) 適合性 + 独立性 1：ボード名に基づいて式 (2) でランキング
- (4) 適合性 + 独立性 2：ボード名に基づいて式 (3) でランキング

各クエリに対して、各ランキング手法の上位 10 件を取得すると、一つのクエリに対して最大 40 件の画像が画像集合として得られる。前節の評価方法と同様に、各クエリに対して得られた画像集合をシャッフルし、画像とクエリを提示しながら、各画像を見てクエリが表す概念としてどれほどそうだと感じるかを評価者に質問し、“強く感じる”を 2 点、“そこそこ感じる”を 1 点、“感じない”を 0 点とする、3 値で評価してもらった。

評価者の回答から、各手法の、各クエリに対する nDCG と全てのクエリに対する nDCG の平均を求めた。nDCG は理想のランキングに対する実際のランキングの良さを表すもので、ランキング手法の評価によく使われている評価指標である。得られた評価結果を図 4 に示す。

図 4 の評価結果を見ると、ほとんどのクエリの場合に対して、3 つの提案手法はベースライン手法より良い精度が得られた。タグは一つの画像に対して類似したタグを何回も付ける場合があまりないことに対し、ピンは一つの画像に対して作成できる上限が存在しなく、いくらでも作成できるので、それぞれのピ

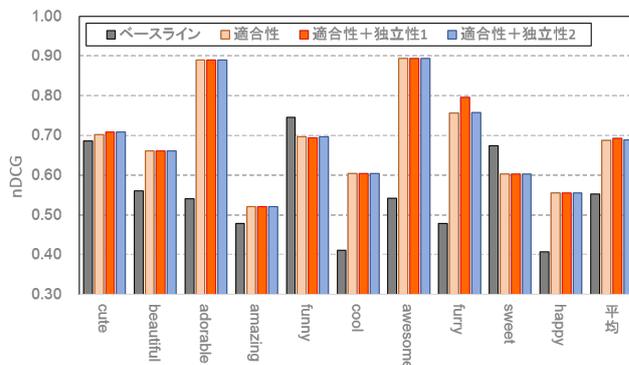


図 4 ランキング手法の有用性に対する評価結果

ンに対応するボード名も無数に存在し、類似したボード名も多く存在することになる。そのため、ベースライン手法より、本研究で提案した、ボード名を用いる数に重みを置いたランキングの精度が高いのではないかとと思われる。従って、この 0.69 の高い nDCG の値が得られたという結果は、提案手法の有用性を示すと同時に、ランキングに用いるデータとしてのボード名の有用性も示している。

3 つの提案手法の間では、精度の差がほとんど見られなかった。実際、実験で得られた各手法の上位 10 件の画像を比較すると、同じ画像が多く得られており、ランキングされた画像集合の間の差があまり見られなかった。それは、実験のクローリングで得られた画像集合では、ピンの数に対してリピンの数が圧倒的に少ないためである。実際の各画像に対する (リピンの数)/(ピンの数) の値を見ていると、約 0.025 の低い値で、これは 40 個のピンが存在すれば、リピンが 1 個含まれることを意味するので、手法 2 と手法 3 のリピンの独立性でスコアを調整する部分があり効果が現れなく、手法 1 とほとんど同じスコアが得られる結果になった。

リピンの数が少ない原因としては、そもそも Pinterest 上にリピンが少ないためであるとも考えられるが、実際は一つのピンに対して数百個のリピンが存在する場合も存在するので、Pinterest 全体を原因として考えるのは早計であると思われる。他の原因としては、クローリングの偏りが考えられる。クローリングというプロセスは、ページ内のリンクを辿りながらデータを集めるものなので、最初のシードとして選択したページによって自然に偏りが発生することになる。本研究で集めた画像集合でも、ボード名に現れる語の頻度に偏りが存在した。つまり、実験で用いた偏った画像集合は、偶然にもリピンの数が少なかったことだけが原因で、提案手法の間に差があまり生まれなかった可能性が高いということである。ピンの独立性を考える手法の良し悪しを考察するためには、今後、クローリングによる偏りを減らした画像集合でもう一度実験を行ってみる必要がある。

6. ま と め

本研究では、ユーザの観点に基づいた画像検索結果のランキングを実現するため、画像のソーシャルキュレーションデータ

を用いて画像検索結果をランキングする手法を提案した。提案手法では、まず与えられたグラフに対応するボード名から観点を抽出し、そのグラフと観点のデータを用い、画像検索結果のランキングしている。印象を表すクエリに対して、タグとボード名の適合率を比較評価し、ボード名がタグより高い値が得られ、ボード名の画像検索における有用性を示した。また、タグを用いてランキングするベースライン手法と、ボード名を用いてランキングする提案手法を比較評価し、0.77, 0.78 という高い nDCG の値が得られ、画像検索結果のランキングにおける提案手法の有用性を示した。

本稿で提案した手法は、5章で述べたように、実験でのクローリングの偏りが原因で、手法の効果を明言できない状態である。また、実験で評価対象とした画像集合は、印象を表すクエリに関連する画像集合で、オブジェクトやイベントまでを含む一般のクエリに対する評価は行っていない。今後は、偏りが小さく、もう少し広い範囲の画像集合を対象として評価を行い、提案手法の有用性を確実に検証し、より高い精度を目指して改善していく予定である。さらに、提案手法をソーシャルキュレーションデータが存在しない画像にも適用できるように、一般の画像とソーシャルキュレーションデータを対応付ける手法の考案にも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面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] Makadia, A., Pavlovic, V. and Kumar, S.: A new baseline for image annotation, *Computer Vision-ECCV 2008*, Springer, pp. 316–329 (2008).
- [2] Csurka, G., Dance, C., Fan, L., Willamowski, J. and Bray, C.: Visual categorization with bags of keypoints, *Workshop on statistical learning in computer vision, ECCV*, Vol. 1, p. 22 (2004).
- [3] Kato, M., Ohshima, H., Oyama, S. and Tanaka, K.: Can Social Tagging Improve Web Image Search?, *Web Information Systems Engineering – WISE 2008*, Vol. 5175, Springer, pp. 235–249 (2008).
- [4] Kimura, A., Ishiguro, K., Yamada, M., Alvarez, A. M., Kataoka, K. and Murasaki, K.: Image context discovery from socially curated contents, *Proceedings of the 21st ACM international conference on Multimedia*, ACM, pp. 565–568 (2013).
- [5] Gilbert, E., Bakhshi, S., Chang, S. and Terveen, L.: “I need to try this”?: a statistical overview of pinterest, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 2427–2436 (2013).
- [6] Linder, R., Snodgrass, C. and Kerne, A.: Everyday ideation: all of my ideas are on pinterest, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 2411–2420 (2014).
- [7] Pedersen, T., Patwardhan, S. and Michelizzi, J.: WordNet::Similarity: measuring the relatedness of concepts, *Demonstration Papers at HLT-NAACL 2004*, pp. 38–41 (2004).