

Web上の地域写真集合における”説得度”のスコアリング(O)

岩木 祐輔[†] アダムヤトフト[†] 田中 克己[†]

[†] 京都大学情報学研究科社会情報学専攻 〒606-8501 京都市左京区吉田本町

E-mail: †{iwaki,adam,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 本論文では Web 上の地域写真集合から「見たらその地域に行きたくなる」ような説得度の高い画像を抽出する方法について論じる．画像の説得度はその写真の地物のよしあしと写真の撮り方のクオリティの二者に依存するものと仮定し，地物のよしあしを Web2.0 コンテンツにおける人気度・典型性から推定し，従来の写真のクオリティ測定の手法とあわせて説得度の定式化を試みる．実験においては，地物のよしあしと写真のクオリティのいずれをどれほど重視するかは地域依存であるが，2 指標の相関関係はほとんどの地域で無相関であることがわかり，さらに，従来のテキストベースの検索と比較して，より高精度で「見て行きたくなる」画像を抽出できることを確認した．

キーワード 写真の説得度，画像検索，信用性評価

How to Calculate the ”Persuasiveness” of Regional Pictures on the Web.

Yusuke IWAKI[†], Adam JATOWT[†], and Katsumi TANAKA[†]

[†] Department of Social Informatics, Graduate School of Informatics, Kyoto University Yoshida-honmachi, Sakyo, Kyoto, 606-8501 Japan

E-mail: †{iwaki,adam,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract In this paper, we discuss how to extract ”persuasive” images in a collection of regional photos on the Web. We presume that the persuasiveness of each photo consists of the subject goodness and the photo quality, and formulate the persuasiveness with approximate subject goodness guessed by Web 2.0 contents and with the previous method of photo quality assesment. In the inspection experiment, we found that the weights of the two indices for persuasiveness depend on the region, and they are almost independent. And we also found our method makes it possible to extract persuasive images more precisely than previous text-based methods.

Key words Picture’s Persuasiveness, Image Retrieval, credibility

1. ま え が き

現在，Web アルバムサービスが広く使われており，代表的なサービスである Flickr や Picasa には，デジタルカメラや携帯電話カメラを用いて撮影された観光などの写真が多数掲載されている．これらの写真情報は実際の体験に基づく情報であるため，これから観光を考えているユーザーなどにとっては有用であると考えられる．しかし，従来の検索においてシステムが求める解は「ユーザーが入力したクエリと関連度の高いコンテンツ」であり，画像検索においては主に周辺テキストをもとに検索をおこなうため，ユーザーが求めるであろう「見て行きたくなる」ようなコンテンツが必ずしも上位に表示されるとは限らない．例えば，屋久島の観光に行きたいというユーザーが図1のように「屋久島 観光」と検索をしても，屋久島観光の記念写真などが，屋久島で人気のある海の写真よりも上位にきている．

そこで，本論文ではユーザーが「見て行きたくなる」画像の抽出する手法を2つの仮定に基づき検討する．まず1点目は，人が説得させられる要素として内面的よさと表面的よさによ



図1 従来の Web アルバム検索における写真の検索結果
屋久島で人気の海の写真などは見られない

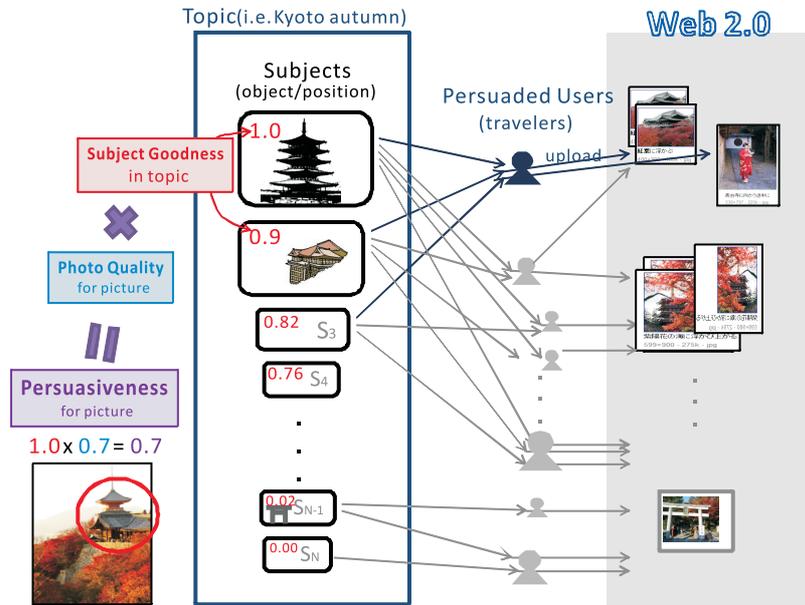


図 2 説得度の定式化のためのモデル

るものであると考える、すなわち、説得度は subject goodness と photo quality で定式化できる、というものである。もう一点は、subject goodness はトピック（先の例では「屋久島 観光」）中の Subject(海、縄文杉など) に対して一意に与えられるとするが、我々は明示的にその値を知ることはできないものとする。以上の仮定で、われわれが提案する手法は、Subject Goodness の値を Web2.0 における実世界での体験コンテンツを集約・分析することで、推定および定式化することである。図 2 にその全体像を示した。

Web2.0 における体験コンテンツの具体的な分析方法として、我々は Persuasive Object もしくは Persuasive Position の 2 点に人が「見ていきたいくなる」画像が集中しやすいと考え、

- 人気がある・典型的な画像を抽出するための従来手法、Visual Rank
 - 人気がある・典型的な撮影位置を抽出するための手法、Position Rank
 - 趣向が偏りすぎているものにペナルティを課すための、Support Rank
- を提案する。これらの全体像は図 3 のようになっている。

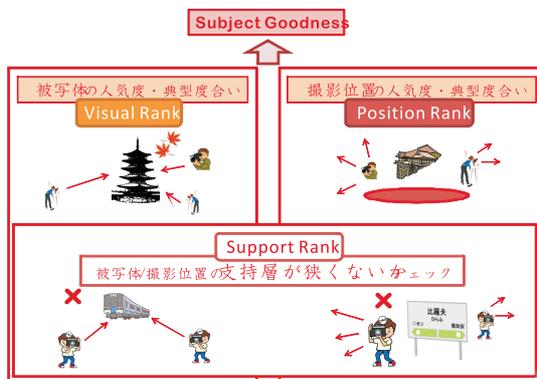


図 3 提案手法の全体図

2. 提案手法

前述のとおり、本研究ではオブジェクトの人気度、あるいは典型度の算出を行うことにするが、そのスコアの計算には Visual Rank の考え方をを用いる。Visual Rank とは、PageRank アルゴリズムの遷移成分を画像類似度に置き換えて計算したスコアである。これによって、画像の画素的に典型的なものが高いスコアを持つような写真集合のスコアリングができる。

2.1 Position Rank

本研究では、Visual Rank は Persuasive Object を抽出するアルゴリズムであると捉え、それとは別に Persuasive Position のスコアを独立に与えるため、同様に写真の撮影位置に対してこのようなアルゴリズムを適用する。すなわち、位置的に近い画像が高い類似度を持つように、PageRank の遷移成分を定め PageRank を計算して、そのスコアを Position Rank とする。具体的な計算方法については以下のようにする。

まず、データのスケールの正規化を行う。 \mathbb{P} それぞれの緯度・経度をそれぞれ標準正規分布 $\mathcal{N}(0, 1)$ に従うよう正規化を行う。

次に、2 点間の近接度を定義する。正規化された 2 点間のユークリッド距離の 2 乗の分布を $\mathbb{D} = \{dist(p_1, p_2)^2 | p_1, p_2 \in \mathbb{P}\}$ とすると、 $\mathbb{D}/2$ は期待値 2 の指数分布に従うことから、2 点 p_1 と p_2 の近接度を以下のように定義する。

$$link(p_1, p_2) = \text{Exp}\left(-\frac{dist(p_1, p_2)}{2}\right)$$

以上のようにして求めた近接度をもとに隣接行列 A を作り、PageRank アルゴリズムに従って計算をし、そのスコアを Position Rank とする。なお、隣接行列 A の (i, j) 成分 $a_{i,j}$ は、

$$a_{i,j} = \begin{cases} link(p_i, p_j) & (i \neq j) \\ 0 & (i = j) \end{cases}$$

のように設定する。

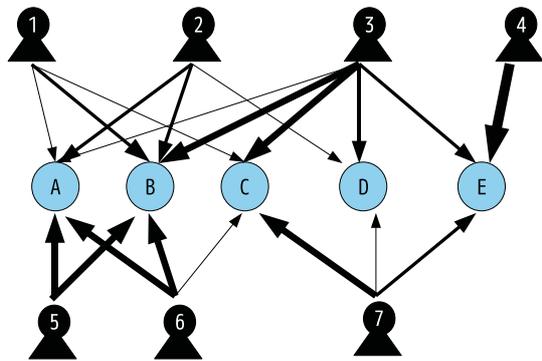


図 4 Support グラフの例

表 1 図 4 の Person Goodness(左) と Support Rank(右)

id	score	id	score
1	0.1915	A	0.3420
2	0.1890	B	0.3422
3	0.1193	C	0.1524
4	0.0259	D	0.1627
5	0.1884	E	0.0006
6	0.1914		
7	0.0944		

2.2 SupportRank

Support Rank は、写真の地物の「支持層の多様性」と「撮影者の趣向の偏りのなさ」をどのくらい有意にいえるか、という指標である。直感的には、趣向の偏りがあり過ぎない人からまんべんなく支持されている地物が高いスコアとなる。例えば、図 4 のような支持グラフであれば、最もまんべんなく支持されている B のスコアが高く、ひどく偏って支持されている E のスコアが低くなる。対象の写真集合 \mathbb{P} に対する詳しい計算手順は以下のとおりである。

まず、前処理として写真集合 \mathbb{P} をトピック別にクラス分けする。たとえば京都の紅葉の写真集合の場合、あらかじめ「清水寺」や「嵐山」などの代表的なトピックを、従来の地域特徴語抽出などの適当な手法で取得しておき、それをもとに「清水寺」に関するクラス、「嵐山」に関するクラス、のように分ける。以上のようにしてトピックに基づいて分類したクラスをそれぞれ C_1, C_2, \dots, C_k とする。

次に、撮影者とクラスタの支持関係を表すための支持行列 S をつくる。 S の (i, j) 成分 $s_{i,j}$ は、クラスタ C_j 中の写真のうち $Person_i$ が撮影したものの枚数とする。さらに、

- C_i を支持している多くの人が C_j も支持している
- $Person_i$ が支持しているものの多くを $Person_j$ も支持している

といった相関性を考慮するために、LSI(Latent Semantic Indexing) を 2 段階で適用する。

$$S_1 = LSI(S)$$

$$S \leftarrow \{LSI(S_1^T)\}^T$$

次に、上で求めた S を用いて、それぞれのクラスタ C_j のよさを計算する。「よいクラスタは広い支持を得ている」という仮

定に基づき、一様分布に基づくクラスタ C_j の支持数の推定値を \hat{s}_j として、クラスタのよさ $ClusterGoodness$ などを以下のように計算する。

$$\chi^2_j = \sum_i \frac{(s_{i,j} - \hat{s}_j)^2}{\hat{s}_j}$$

$$ClusterGoodness_j = \frac{\chi^2_j}{\chi^2(k-1)_{0.05}}$$

$$ClusterBadness_j = \frac{1}{ClusterGoodness(C_j)}$$

($\chi^2(k-1)_\alpha$ は自由度 $(k-1)$ の危険度 α に対する χ^2 値)

さらに、 $ClusterGoodness$ と $ClusterBadness$ を用いて、撮影者の趣向の偏りを計算する。「よい撮影者は、よいクラスタを(よくないクラスタに比べてより多く)まんべんなく支持する」という仮定に基づき、

$$\hat{s}G_{i,j} = \frac{ClusterGoodness_j}{\sum_j ClusterGoodness_j} n_i$$

$$\hat{s}B_{i,j} = \frac{ClusterBadness_j}{\sum_j ClusterBadness_j} n_i$$

(n_i は $Person_i$ の撮影総数 $\sum_j s_{i,j}$)

のように総支持数をそれぞれ比例配分した推定値を用いて、撮影者のよさ $PersonGoodness$ などを以下のように計算する。

$$\chi^2_{G_i} = \sum_j \frac{(s_{i,j} - \hat{s}G_{i,j})^2}{\hat{s}G_{i,j}}, \quad \chi^2_{B_i} = \sum_j \frac{(s_{i,j} - \hat{s}B_{i,j})^2}{\hat{s}B_{i,j}}$$

$$F_i = \frac{\chi^2_{G_i}}{\chi^2_{B_i}}$$

$$PersonGoodness_i = \frac{F_i}{F_{k-1}^{k-1}(0.05)}$$

$$PersonBadness_i = \frac{1}{PersonGoodness(C_j)}$$

($F_{k-1}^{k-1}(\alpha)$ は自由度 $(k-1)$, $(k-1)$ の危険度 α に対する F 値)

さらに、 $PersonGoodness$ と $PersonBadness$ を用いて、クラスタのよさを再度求める。これが Support Rank である。「本当によいクラスタは、よい撮影者らによって(よくない撮影者に比べてより多く)まんべんなく支持される」という仮定に基づき、 $PersonGoodness$ と同様、以下のように計算を行う。

$$\hat{s}G_{i,j} = \frac{PersonGoodness_i}{\sum_i PersonGoodness_i} \cdot \#(C_j)$$

$$\hat{s}B_{i,j} = \frac{PersonBadness_i}{\sum_i PersonBadness_i} \cdot \#(C_j)$$

($\#(C_j)$ はクラスタ C_j の写真総数 $\sum_i s_{i,j}$)

$$\chi^2_{G_j} = \sum_i \frac{(s_{i,j} - \hat{s}G_{i,j})^2}{\hat{s}G_{i,j}}, \quad \chi^2_{B_j} = \sum_i \frac{(s_{i,j} - \hat{s}B_{i,j})^2}{\hat{s}B_{i,j}}$$

$$SupportRank(C_j) = \frac{\chi^2_{G_j}}{\chi^2_{B_j}}$$

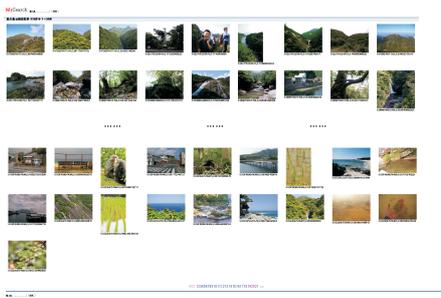


図5 「屋久島」の検索例

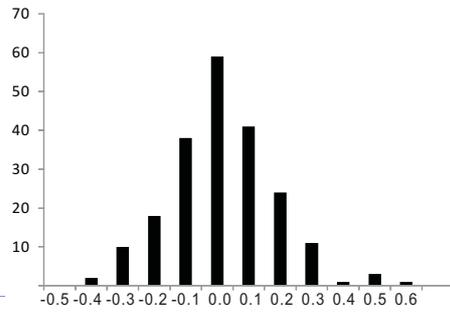


図6 Visual/Position Rank の相関係数の分布

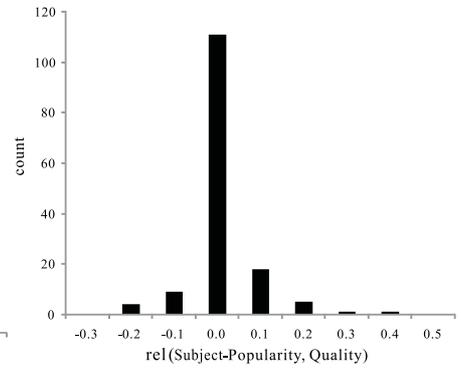


図7 Subject Goodness と Photo Quality の相関係数の分布

図4の支持グラフをもとに Person Goodness と Support Rank を計算して正規化すると、表1のようになる。

3. 評価

これまでに述べた手法を実装して検索結果の観察および精度評価を行った。ベースライン手法としては、Picasa Web アルバムによるキーワード検索を用いた。

まず、「屋久島」や「久留米」のように、名物が既知であるいくつかの地名について観察を行ったところ、本研究の提案システムでは図5のように、上位には人気なものや典型的なものがランクインしており、従来の検索と比較すると見ていきたくなくなるコンテンツが上位にきている可能性が高いといえる。ただし予備実験において、Subject Goodness と Photo Quality のマージの割合は地域ごとに異なることを確認しているため、ここでは適合率再現率評価などの定量的な評価は行わなかった。なお、Persuasive Position と Persuasive Object の相関係数および Subject Goodness と Photo Quality との相関係数は図6・7のとおりであり、いずれもほぼ無相関であることを確認した。

また、表2のようにいくつかのクエリにおいて上位20件の検索結果の精度を比較したところ、多くの例で本研究の提案手

法がベースライン手法を上回った。

広島など精度が下がっている例も見られるが、これは人の属性を考慮していないことによる不適合である。具体的には、住民運動会や競技大会など多くの人が大量の写真を撮影してアップロードしている際に、それらが当該地域において人気・典型的であると誤判定されている。現在のモデルにおいては、単純のため、人は「実世界をキャプチャしてアップロードするマシン」と捉えているため、今後は人の属性も考慮した手法を提案していく必要がある。

4. まとめと今後の課題

本論文では、説得度の高い「見て行きたくなる」画像を抽出する手法について提案した。説得の要素でありながら現状の検索では保障されていない写真の地物の客観性を補うため、Position Rank と Support Rank の2つの指標を提案した。従来の周辺テキストに基づく方法と比較して、これらの2つの指標を導入することで精度が向上することを実データを用いて検証した。

今後の課題としては、画像の説得度を定める要因としてこれらの2点が最適であるかの検証、および位置情報や撮影者情報が付加されていない一般画像への応用の2点が挙げられる。

表2 Precision@20

地名	提案手法	ベースライン
稚内	0.50	0.35
札幌	0.55	0.30
帯広	0.30	0.25
仙台	0.20	0.30
水戸	0.60	0.05
伊豆	0.25	0.45
富山	0.40	0.35
高山	0.45	0.35
和歌山	0.45	0.30
潮岬	0.70	0.35
松江	0.65	0.10
広島	0.15	0.20
長崎	0.40	0.50
石垣島	0.35	0.45
屋久島	0.90	0.05