

画像検索におけるクエリ多言語化の評価

高橋 有礼

菱山 玲子

早稲田大学理工学術院 創造理工学部経営システム工学科

{arinori@ruri, reiko}@waseda.jp

1 研究背景と目的

Web上で蓄積されるマルチメディアコンテンツの増加に伴い、これらの情報取得のための様々な手法が提案されている。その手法のひとつとして、タグ付けされたコンテンツ内容の記述を活用してマルチメディア情報を取得する手法がある。しかし、この方法は付与されたタグが適切でない場合には有効に機能せず、これがマルチメディア情報取得の新たな課題として挙げられている [1]。また、コンテンツ内容の記述はもともと多言語で付与されている。よって、単言語による検索クエリやタグ情報を利用するだけでは、コンテンツの適切な取得が困難と予測される。この点で、検索クエリとして情報の詳細さを多言語で記述しきれない場合や、ユーザ自身が検索クエリを適切に書き起こせない場合が想定される。

そこで、本研究ではマルチメディアコンテンツのひとつである静止画像を対象として、機械翻訳サービスを利用してユーザの検索クエリを多言語化して生成する。この多言語化した検索クエリの情報を活かし、静止画像に含まれるメタデータを効果的に利用しながら正解画像に至るまでのプロセスを改良することで、ユーザ意図に沿った適切な画像取得を実現する。

2 関連研究

言語横断的な検索手法に関し、Kristenら [2] は、検索クエリと文章の双方を翻訳することで検索精度の向上を提案している。また、Paulら [3] は、クエリ翻訳とドキュメント翻訳を比較してどちらがよりよい結果を返すのかを研究している。実験はドキュメントに関しては英語以外のものを英語へ翻訳すること、英語を仏語、独語、伊語、西語へ翻訳することで行われており、クエリに関しては他の言語を英語に翻訳することのみが行われている。実験の結果、クエリを翻訳したほうがドキュメントを翻訳するよりも適合率が高くなった。しかし、これらの研究は言語横断的な多言語クエリの組み合わせによる効果が検証されていない。

画像検索を効果的に行う方法として、Zakariaら [1] はメタデータに依存することなく、ユーザ主体のタグ付けや地理参照メタデータによって生成された集合的な知識と組み合わせることで、コンテンツ内容記述の充実化を図った。しかし、これらにはユーザの検索意図を直に反映させた検索を考慮したものではなく、付与タグの多言語化も検討されていない。

3 提案

本研究では、ユーザが期待する画像の取得を実現するため、以下の方法を提案する。まず、正解画像を入手する過程に、中島ら [4] の差異増幅型フィードバックを適用し、ユーザの意図を反映した画像検索を実現する。このフィードバック手法を効果的に利用するためには、検索対象となり得る画像を幅広く入手しておく、ユーザの適正画像の選択がより適切なものとなるよう導くことが必要である。すなわち、フィードバック時にユーザが適合画像として選択する画像群の候補集合を拡張することが効果的である。そこで、この候補集合の拡張を行うため、画像検索前に機械翻訳サービスを用いてユーザ使用言語で記述された検索クエリを多言語に翻訳すると同時に、メタデータとして多言語情報を活用しながら、効果的にこれらを組み合わせる。これにより、コンテンツ内容の記述を多言語で補填しながら、検索対象とするタグ候補と量を増大することで、検索精度を向上させる。本研究の流れを図1に示し、各処理機能の詳細を以下に記述する。

データベース部

データベース部は、検索対象の画像とそれに付与されたタグ、そしてタグを後述するメタデータ処理部で処理した結果を格納している。類似度計算時にはこのデータベースに格納されている画像すべてに対して、類似度計算を行う。

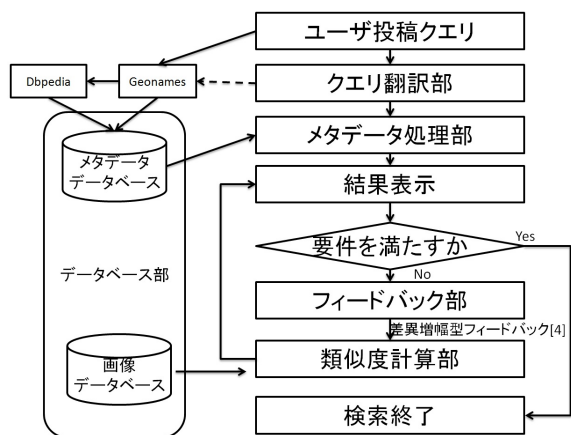


図 1. 提案フロー

メタデータ処理部

メタデータ処理部では、画像に付与されたタグデータをもとに、既存の Web サービス連携により集合的に地理参照メタデータと組み合わせ、コンテンツ内容の記述の充実化を図る。具体的な処理は以下のとおりである。まず、Geonames[5]を用いて、対象を検索する。Geonamesとは、全世界の地理データを検索可能としている Web サービスである。これにより対象を、地名・建造物名とそれ以外の 2 種類に分類する。また、DBpedia[6]により、Geonames で地名・建造物名かどうかを判別したものと同一のものを取得するため、Geonames が独自に割り振っている GeonameID を取得する。その後、地名・建造物名に対しては、DBpedia から関連する地名を取得し、これをメタデータとして追加する。DBpediaとは、Wikipedia から情報を抽出して Linked Open Data として公開している Web サービスである。

クエリ翻訳部

クエリ翻訳部では、検索開始前にユーザー使用言語で記述されたクエリを、検索対象の現地言語や英語に翻訳し追加する。これにより、新たに多言語化された検索クエリを生成する。

フィードバック部

フィードバック部はユーザーの検索意図を反映した検索を実現する機能であり、中島ら [4] により提案された差異増幅型フィードバックを適用する。この方法は、ユーザーが選択したコンテンツと選択しなかったコンテンツの差異を利用し、ユーザーの意図をより強く反映させるものである。ここで用いる関数 f_{DA} は正事例 S_{pi} 、正事例の周辺画像 $neighbor(S_p)$ 、正事例の周辺画像数

$N_{S_{pi}}$ で定義される。 α は差異増幅の増幅係数であり、 ϵ は差異を増幅させるかの閾値である。 S_{pi} に基づく次の検索質問 Q_{i+1} を以下の式 (1) と定義する。

$$Q_{i+1} = (1-\beta) \times Q_i + \beta \times f_{DA}(S_p, neighbor(S_p)) \quad (1)$$

ただし

$$f_{DA}(S_{pi}, neighbor(S_{pi}))$$

$$= \begin{cases} S_{pi} + \alpha \times \frac{\sum_{i \in neighbor(S_{pi})} (S_{pi} - S)}{N_{S_{pi}}} & (|\frac{\sum_{i \in neighbor(S_{pi})} (S_{pi} - S)}{N_{S_{pi}}}| \geq \epsilon) \\ S_{pi} & (|\frac{\sum_{i \in neighbor(S_{pi})} (S_{pi} - S)}{N_{S_{pi}}}| < \epsilon) \end{cases}$$

類似度計算部

類似度計算部では、ユーザーが選択した画像とデータベース内の画像の類似度を 64 色に減色した状態で計算する。減色は画像の各ピクセルに対して、RGB の各成分を 4 等分し中央値に置き換えることで行う。そして減色後に 64 次元のカラーヒストグラムを作成する。カラーヒストグラム生成後は類似度を計算する。類似度計算には HistogramIntersection[7]を用いる。この式 (2) では H_1 が画像クエリのヒストグラムであり、 H_2 は比較対象のヒストグラムである。この二つのヒストグラムの同次元の要素 i を比べ、値が小さい方を足し合わせていくことで 2 つの画像の類似度を求める。ただし画像の大きさにヒストグラムの要素数は依存するため、正規化を行っている。この類似度が高い順に画像を整頓し、その結果をユーザーに対して示す。

$$\frac{\sum_{i=0}^{63} \min(H_1[i], H_2[i])}{\sum_{i=0}^{63} H_1[i]} \quad (2)$$

4 実験

評価実験は、各国語によるクエリによる検索結果比較 (実験 1) と、各国語によるクエリの組み合わせによる検索結果比較 (実験 2) により行った。

4.1 実験 1 の設定

実験 1 は、次の 2 種の実験からなる。実験 1.1 では検索対象を「見上げた角度、赤色、エッフェル塔」を満たすものとし、対象の画像は Flickr[8] からエッフェル塔関連のタグがつけられた画像 785 枚を取得した。実験 1.2 では検索対象を「ロゴを読むことができる、赤色、アリアンツアリーナ」を満たすものとし、対象の画像は Flickr[8] からアリアンツアリーナ関連のタグがつけられた画像 701 枚を取得した。また取得した画像に付与されたタグも加えることでデータベースを構

築した。検索成功条件は適合率が1/3以上となった時とし、検索失敗条件は適合率が0の時とした。画像の選択基準は検索結果最上位の画像とした。但し、同一画像が2回続いた場合は検索結果2番目の画像としている。実験は、以下の単言語クエリで検索を行う。

実験 1.1

- Tour+Eiffel+Rouge(仏)
- Eiffel+Tower+Red(英)
- エッフェル塔+赤(日)

実験 1.2

- Allianz+Arena+Rot(独)
- Allianz+Arena+Red(英)
- アリアンツアリーナ+赤(日)

4.2 実験 2 の設定

実験 2 では実験 1 のクエリを現地言語に翻訳したものと、英語に翻訳したものをクエリに追加した2種類3パターンの多言語クエリで検索を行う。実験は、以下の多言語クエリで検索を行う。

実験 2.1

- Eiffel+Tower+Red+Tour+Eiffel+Rouge(英+仏)
- エッフェル塔+赤+Tour+Eiffel+Rouge(日+仏)
- エッフェル塔+赤+Eiffel+Tower+Red(日+英)

実験 2.2

- Allianz+Arena+Red+Allianz+Arena+Rot(英+独)
- アリアンツアリーナ+赤+Allianz+Arena+Rot(日+独)
- アリアンツアリーナ+赤+Allianz+Arena+Red(日+英)

5 結果と考察

5.1 結果

各実験の、開始時と終了時の適合率を表 1、表 2、表 3、表 4 に示し、検索結果を図 2、図 3、図 4、図 5、図 6 に示す。開始時結果とは図 1 の提案フローにおいてメタデータ検索後の初回の結果であり、終了時とは検索成功/失敗条件を満たした際の結果である。また開始時適合率とは開始時結果の適合率であり、終了時適合率とは終了時結果の適合率である。

	開始時適合率	終了時適合率
仏	27	47
英	10	33
日	0	0

	開始時適合率	終了時適合率
独	27	57
英	30	53
日	0	0

	開始時適合率	終了時適合率
英+仏	27	47
日+仏	27	47
日+英	13	43

	開始時適合率	終了時適合率
英+独	27	53
日+独	30	53
日+英	36	36

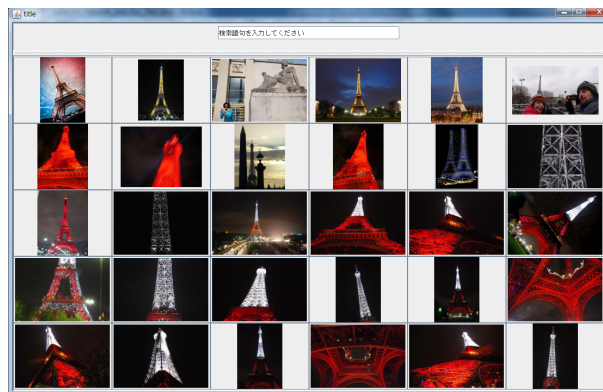


図 2. Tour+Eiffel+Rouge 開始時結果

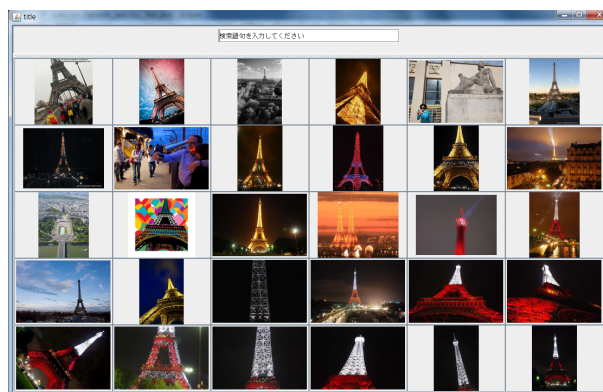


図 3. Eiffel+Tower+Red 開始時結果

5.2 考察

表 1、表 2 では現地言語では開始時で多くの適合画像が入手できるが、日本語では1件も出力されなかつ

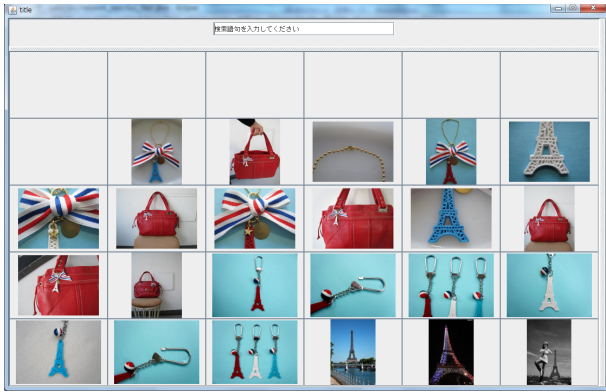


図 4. エッフェル塔+赤 開始時結果

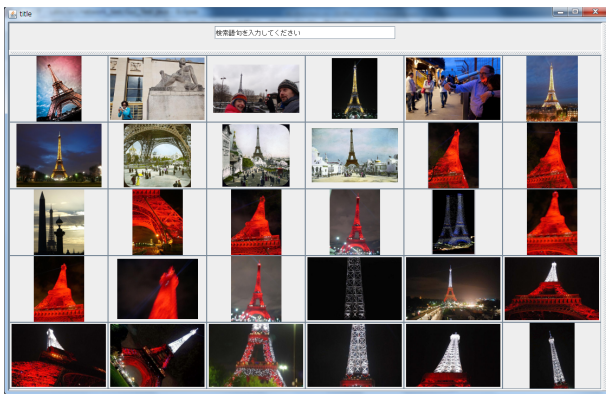


図 5. エッフェル塔+赤+Eiffel+Tour+Rouge 開始時結果

た。このことから、検索クエリの言語は開始時の検索結果に影響を及ぼすことがわかる。表3、表4ではユーザのクエリを現地言語に翻訳し追加することで、追加前は適合画像が1枚もなかった言語でも追加後は多数の適合画像が入手できている。実験2.2において英と独が適合率に差が余り見られないのは、対象の表記名が英、独で変わらないため言語による差があまり出なかったからである。以上から、適合画像は、ユーザの使用言語でタグ付けされたものよりも検索対象の現地の言語でタグ付けされたものに多く存在する。そしてクエリに現地の言語を翻訳したものを追加することで、開始時に多くの適合画像を得ることができる。

6 まとめと今後の課題

本研究では、多言語クエリを利用し、画像に含まれる情報及びメタデータの効果的利用から正解画像に至るまでのプロセスを改良することで、より適切な画像の入手を実現した。検索クエリ生成時の言語選択は、検索結果に大きな影響を及ぼし、ユーザ意図に沿った検索の実現にも影響を与える。対象画像の地理的情報を効果的に利用し検索クエリの言語を選択することで、

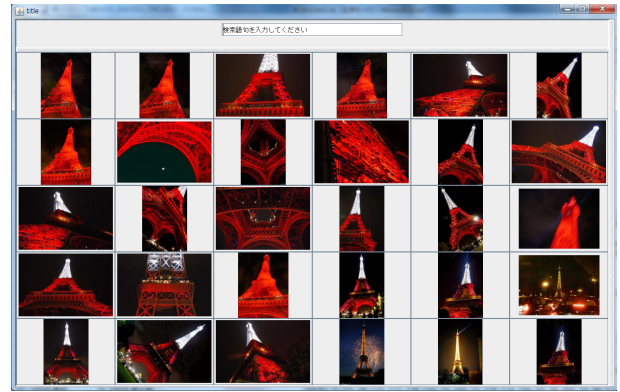


図 6. エッフェル塔+赤+Eiffel+Tour+Rouge 終了時結果

ユーザによるフィードバックを支援し、適合画像を容易に取得することができることがわかった。今後の課題は、更に多様な画像による評価を行うと共に、クエリが複雑で翻訳が困難な場合を考慮した方法を検討することである。

謝辞

本研究の一部は、早稲田大学特定課題研究助成費(2013B-113)の補助を受けた。

参考文献

- [1] Lailatul Qadri Zakaria, Wendy Hall, and Paul Lewis: Modelling image semantic descriptions from web 2.0 documents using a hybrid approach, *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services*, pp.306-312, 2009.
- [2] Parton Kristen., et al. : Simultaneous multilingual search for translingual information retrieval, *Proceedings of the 17th ACM conference on Information and knowledge management*, pp.719-728, 2008.
- [3] Clough Paul: Caption and query translation for cross-language image retrieval. *Multilingual Information Access for Text, Speech and Images*. Springer Berlin Heidelberg, pp.614-625, 2005.
- [4] 中島伸介, 木下真一, 田中克己: 差異増幅型適合フィードバックに基づく画像データベース検索, *電子情報通信学会論文誌 D-1*, Vol.87, No.2, pp164-174, 2004
- [5] Geonames, URL: <http://ws.geonames.org/>, (2014年1月9日アクセス)
- [6] DBpedia, URL: <http://dbpedia.org/>, (2014年1月9日アクセス)
- [7] Swain, Michael J., and Dana H. Ballard: Color indexing, *International journal of computer vision*, Vol.7, No.1, pp.11-32, 1991.
- [8] Flickr, URL: <http://www.flickr.com/>, (2013年12月10日アクセス)