

F-11

画像とタグによるクロスメディア適合性フィードバック

Cross-Media Relevance Feedback Using Images and Tags

杉山 裕樹[†]
Yuki Sugiyama

加藤 誠[†]
Makoto P. Kato

大島 裕明[†]
Hiroaki Ohshima

田中 克己[†]
Katsumi Tanaka

1. はじめに

ブロードバンド・インターネットの進展に伴い、Web のリッチメディア化が進展している。Web 情報としては既にテキストのみならず、画像・動画・音声など様々なメディアが活用されており、このようなデータを含んだオブジェクトを検索する機会も増えてきている。

Web 情報検索は、そもそもユーザが知らない情報を検索によって求めることが主であるが、未知の情報に対して、適切な検索クエリをユーザが指定することは非常に困難なことである。

例えば現在の画像検索では、画像の周辺テキストや画像ファイル名などの情報をもとに、キーワードによる検索を行うのが一般的であるが、ある犬種の画像を検索したい場合に、外観は思い出せるがその犬種名を思い出せないといった状況では、ただ単に「犬」や「茶色い犬」、「小型犬」のように曖昧で不完全なクエリしか入力できない。

このような不完全なクエリでは、ユーザの検索意図に沿った結果を得ることは困難であるため、クエリを修正したり、新たなキーワードを付け加えたりして再検索を行わなければならない。

ユーザのクエリ修正を助ける手法として代表的なものに、適合性フィードバック[1]がある。これは、ユーザが検索対象に対して行った正誤判定をもとに、その検索対象の特徴ベクトルを利用してクエリベクトルを修正する手法であるが、通常の適合性フィードバックでは、検索対象の特徴ベクトルと同一の特徴空間にあるクエリにしかフィードバックすることができない。

しかし、テキスト・画像・動画などを含むマルチメディアオブジェクトは、様々な特徴量の表し方があり、ユーザが行ったフィードバックの意図も多様に存在する。例えば、説明文と画像を含むデータに対して正のフィードバックが与えられた場合、説明文と画像のどちらがユーザにとって正解であるのか、また、両方とも正解であるのかを考慮してクエリ修正を行わなければならない。また、正誤判定の対象となるメディア、そしてその正誤判定の範囲も考慮する必要がある。例えば、あるデータに含まれるテキスト全体を指定したい場合やその中の 1 単語のみを指定したい場合、複数の画像のうちのある画像全体の色合いを指定したい場合や茶色という色だけを指定したい場合など、様々な状況が考えられる。

本研究では、複数のメディアで表現されたオブジェクトを検索する際に、柔軟な適合性フィードバックを行うため、メディア毎の特徴量とそれに対応するクエリを導入し、これらの複数の特徴量を交差的に利用する、クロスメディア適合性フィードバック検索を提案する。

これにより、図 1 に示すように、出力された検索結果から、ユーザはオブジェクト全体や各メディア毎の特徴ベクトル、特徴ベクトルの次元など、自由に正誤判定の対象レベルを選択することが可能となる。

クロスメディア適合性フィードバックは、メディア毎の特徴量を異メディアの特徴空間へ変換し、適合性フィードバックを異なる特徴量へ相互伝播させ、ユーザの検索意図をクエリにくまなく伝えることが可能な検索モデルを提案する。

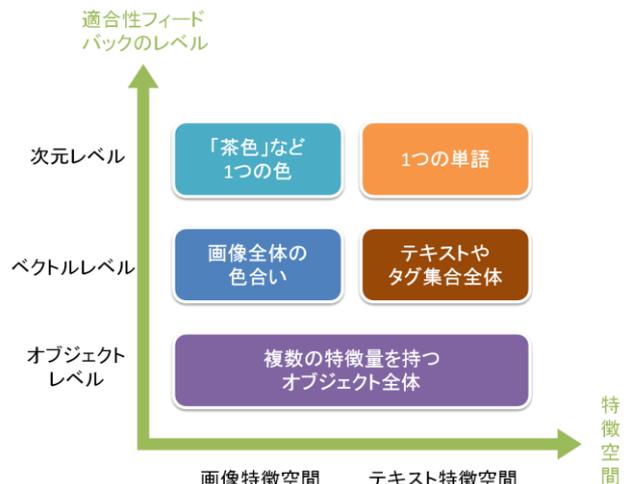


図 1 適合性フィードバックのレベル



図 2 クロスメディア適合性フィードバック

[†] 京都大学大学院 情報学研究科 社会情報学専攻

図2に示すように、通常の適合性フィードバックでは画像から画像、テキストからテキストと独立にしかフィードバックできなかったが、提案するクロスメディア適合性フィードバックによって、画像からテキスト、テキストから画像と特徴空間を跨いだフィードバックが可能となる。

本稿では Flickr[2]において、画像特徴量 Color Coherence Vector[3]と画像に付与されたタグ集合を利用したときの、フィードバック手法など、本研究の検索モデルの適用例を示す。

2. 関連研究

適合性フィードバック[1]とはベクトル空間モデルの情報検索において、適合文書や不適合文書の特徴ベクトルからクエリを再定式化する手法である。Rocchio は以下の式により、クエリベクトルの修正を行っている。

$$\mathbf{q}(t+1) = \alpha \mathbf{q}(t) + \frac{\beta}{|D_r|} \sum_{\mathbf{d} \in D_r} \mathbf{d} - \frac{\gamma}{|D_{nr}|} \sum_{\mathbf{d} \in D_{nr}} \mathbf{d}$$

ただし、 $\mathbf{q}(t)$ は時刻 t におけるクエリベクトル、 \mathbf{d} は文書の特徴ベクトル、 D_r, D_{nr} はそれぞれ適合、不適合文書集合であり、 α, β, γ は元クエリや各フィードバックの重みを定める定数である。

この式は全適合文書の平均特徴ベクトルから全不適合文書の平均特徴ベクトルを引いたものが、最適化されたクエリベクトルであるという考えに基づいて、クエリを段階的に修正していくものである。

Rocchio の適合性フィードバックはベクトル空間モデルの情報検索において有効であることが Salton ら[4]によって示されている。また、Rui ら[5]は画像検索において、画像特徴量に関して適合性フィードバックを行うことで、有効性を示している。

しかし、Rocchio の式はクエリベクトルに文書の特徴量をそのまま足し合わせているため、同一の特徴空間の中でしかフィードバックを行うことができない。本研究では複数のメディアを持ったオブジェクトに対して適合性フィードバックを適用する際に、特徴空間を跨いだフィードバックを行う手法を提案する。

3. 検索モデル

3.1 オブジェクトと特徴量

検索対象となるオブジェクト $o_k (k=1, \dots, n)$ には、特徴ベクトルで表現できる m 種類の特徴量がある。各特徴ベクトル空間を V_1, V_2, \dots, V_m とし、オブジェクトを o_k 、オブジェクトを各特徴空間の特徴量へ変換する関数を $f_i(o_k)$ とする。

3.2 クエリ

検索におけるクエリは、各特徴ベクトル空間 V_i の特徴ベクトル $\mathbf{q}_i(t)$ の組で表現される。すなわち、時刻 t におけるクエリは

$$\mathbf{Q}(t) = (\mathbf{q}_1(t), \mathbf{q}_2(t), \dots, \mathbf{q}_m(t))$$

$$\mathbf{q}_i(t) \in V_i$$

となる。

3.3 クエリ修正

クロスメディア適合性フィードバックでは、ある特徴ベクトル空間 V_i に関連するユーザのフィードバックは、別のベクトル空間 V_j に属する特徴ベクトル $\mathbf{q}_j(t+1)$ にも修正を加える。このメディア間の相互修正が既存の適合性フィードバックと大きく異なる点である。クエリ修正は Rocchio の適合性フィードバックと同じように、前の時刻のクエリから正例の総和を足し、負例の総和を引いた形で表される。すなわち、時刻 t におけるクエリ $\mathbf{Q}(t)$ は、ユーザからのフィードバック操作をもとに次式のように修正される。

$$\mathbf{Q}(t+1) = (\mathbf{q}_1(t+1), \dots, \mathbf{q}_m(t+1))$$

$$\mathbf{q}_j(t+1) = \mathbf{q}_j(t) + \sum_{i=1}^m M_{ij}(\mathbf{v}_i^+) - \sum_{i=1}^m M_{ij}(\mathbf{v}_i^-)$$

ただし、 \mathbf{v}_i^+ はユーザが指定した正例に対応する空間 V_i のベクトルであり、 \mathbf{v}_i^- は負例に対応するベクトルである。また、 M_{ij} はベクトル空間 V_i から V_j への写像関数とする。この $\mathbf{v}_i^+, \mathbf{v}_i^-$ 及び M_{ij} はユーザのフィードバック操作によって決定される。例えば、ユーザがオブジェクト o_k 全体に対して正のフィードバックを与えた場合、 \mathbf{v}_i^+ としてそのオブジェクトの特徴ベクトル空間 V_i による表現である $f_i(o_k)$ が考えられ、 M_{ij} としては特に変換を行わず単位行列を掛けるだけといった関数が考えられる。しかし、これらは用いる特徴量表現の種類に大きく依存するため、4節でその具体的な例を与える。

3.4 ランキング

情報検索において、あるクエリに対する検索対象データの順位はランキング関数によって決定される。ランキング関数はクエリと検索対象データから実数値への関数であり、この値が大きい順に検索対象データが順位付けされる。クロスメディア適合性フィードバックでは、ランキング関数は各特徴ベクトル空間 V_i に対応するクエリ \mathbf{q}_i とそのベクトル空間で表されたオブジェクト o_k の類似度によって計算される。すなわち、時刻 t のクエリに対するオブジェクトのランキングは、次式によって定義されるランキング関数によって行う。

$$\text{Rank}(o_k, \mathbf{Q}(t))$$

$$= \Phi(\text{Sim}_1(f_1(o_k), \mathbf{q}_1(t)), \dots, \text{Sim}_m(f_m(o_k), \mathbf{q}_m(t)))$$

ただし、 Sim_i は空間 V_i における、ベクトルの類似度であり、 Φ は m 個の類似度を1つに組み合わせる関数である。つまり、

$$\text{Sim}_i : V_i \times V_i \rightarrow [0,1]$$

$$\Phi : [0,1]^m \rightarrow [0,1]$$

となる。 Sim_i にはコサイン類似度やユークリッド距離が、 Φ には総和や相乗などが考えられる。しかし、特徴ベク

トル空間の種類によって効果的な Sim_i は異なるため、また、一般化のためここでは明確に定義は与えず、実装において議論を行う。

4. 実装

4.1 オブジェクトと特徴量

検索対象として Flickr の画像ページを取り上げる。また、その特徴量としては画像特徴量である Color Coherence Vector(CCV)[3]と、テキスト特徴量のタグ集合を利用した。

CCV は色ヒストグラム的一种であるが、似ている色が密になっているか疎になっているかも表すことができ、空間的な情報も捉えることができる。

また、タグ集合は複数のキーワードからなり、その特徴量はキーワードを次元としたベクトルで表され、画像にタグとして付与されているキーワードの次元を 1、残りを 0 としたものである。

4.2 フィードバック操作

ユーザのフィードバック操作としては、以下の 3 つを考える。

- (i) オブジェクト選択：
オブジェクト o_k を選択する。
- (ii) 特徴ベクトル選択：
あるオブジェクトの特徴ベクトル $f_i(o_k)$ を選択する。
- (iii) 次元選択：
ある特徴ベクトル空間 V_i の次元を選択する。

これらのフィードバックは、ユーザのフィードバック意図など異なった性質を持っているため、適切なフィードバック手法も異なる。例えば、オブジェクトを選択した場合は、全ての特徴量を選択したと考えられるため、メディア間の伝播は行わずとも、全てのクエリを修正することができる。一方、次元を選択した場合には、その次元に値を持つオブジェクトを得るなどといった操作が必要となる。また、ベクトルを選択した場合は、ユーザが選んだ特徴量をはっきりしているため、類似オブジェクトを取ることが容易であるといった特徴がある。これらのフィードバックの詳しい定義は次項で述べる。

4.3 写像関数

フィードバックの写像関数はフィードバック方法によって異なる。前項で設定した 3 種類のフィードバックに対し、それぞれ定義を与える。

- (i) オブジェクトレベルの適合性フィードバック
オブジェクトを選んだ場合、そのオブジェクトは全ての特徴量を持っているため、従来の適合性フィードバックと同様に、選択したオブジェクトの特徴量をそのままクエリベクトルに足し引きすればよい。すなわち、オブジェクト o_k が選択された場合、写像関数 M_{ij} とベクトル \mathbf{v}_i は以下のように定義できる。

$$M_{ij}(\mathbf{v}_i) = \begin{cases} 0 & (i \neq j) \\ \mathbf{v}_i & (i = j) \end{cases}$$

$$\mathbf{v}_i = f_i(o_k)$$

- (ii) ベクトルレベルの適合性フィードバック

ベクトルを選んだ場合、そのベクトルと特徴量が類似したオブジェクトを取得することができる。この類似オブジェクトの特徴量も利用することで、特徴量の補填が可能となる。

ここで、ある特徴量に関して、全てのオブジェクトの特徴量を並べた行列 \mathbf{A}_j を考える。

$$\mathbf{A}_j = (f_j(o_1) \quad f_j(o_2) \quad \cdots \quad f_j(o_n))$$

この行列を用いることで、各オブジェクトの特徴量を、選択したベクトルとの類似度に合わせてフィードバックを行う写像関数 M_{ij} が定義できる。

$$M_{ij}(\mathbf{v}_i) = \begin{cases} \mathbf{A}_j \begin{pmatrix} \text{Sim}(\mathbf{v}_i, f_i(o_1)) \\ \vdots \\ \text{Sim}(\mathbf{v}_i, f_i(o_n)) \end{pmatrix} & (i \neq j) \\ \mathbf{v}_i & (i = j) \end{cases}$$

ベクトル \mathbf{v}_i はユーザによって選択された特徴ベクトル $f_i(o_k)$ である。

- (iii) 次元レベルの適合性フィードバック

行列 \mathbf{A}_j の転置行列を用いることで、選択した次元に値を持つオブジェクトを得ることができる。これらのオブジェクトが持つ特徴量を利用することで、次元によるフィードバックを定義できる。ただし、特徴空間が変わらない場合は、ベクトルを変化させる必要がない。従って、写像関数 M_{ij} は次のようになる。

$$M_{ij}(\mathbf{v}_i) = \begin{cases} \mathbf{A}_j \mathbf{A}_i^T \mathbf{v}_i & (i \neq j) \\ \mathbf{v}_i & (i = j) \end{cases}$$

ベクトル \mathbf{v}_i は空間 V_i のベクトルのうち、ユーザが選んだ次元が 1、残りの次元が 0 のベクトルである。

4.4 類似度

画像特徴量 CCV の類似度は、2 ベクトル \mathbf{a}, \mathbf{b} 間の距離 $\text{dist}(\mathbf{a}, \mathbf{b})$ を用いて次のように定義する。

$$\text{dist}(\mathbf{a}, \mathbf{b}) = \sum_i |a_i - b_i| \in [0, \infty)$$

$$\text{Sim}_{\text{ccv}}(\mathbf{a}, \mathbf{b}) = \exp(-\text{dist}(\mathbf{a}, \mathbf{b})) \in (0, 1]$$

タグ集合の類似度は、2 ベクトル \mathbf{a}, \mathbf{b} のコサイン類似度を用いる。つまり、以下のように定義する。

$$\text{Sim}_{\text{tag}}(\mathbf{a}, \mathbf{b}) = \frac{1}{2} \left(1 + \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \right) \in [0, 1]$$

ただし、 $\mathbf{a} \cdot \mathbf{b}$ はベクトルの内積であり、 $\|\mathbf{a}\|, \|\mathbf{b}\|$ はベクトルのユークリッドノルムである。

4.5 ランキング関数

画像の出力順位を決めるランキング関数における関数 Φ は、2 つの関連度 $\text{Sim}_{\text{ccv}}, \text{Sim}_{\text{tag}}$ が共に大きいときの大きな値を取るよう、次のように定義する。

$$\Phi(\text{Sim}_{\text{ccv}}, \text{Sim}_{\text{tag}}) = \text{Sim}_{\text{ccv}} \cdot \text{Sim}_{\text{tag}}$$

この式によって計算された $\text{Rank}(o_k, \mathbf{Q}(t))$ の値の大きいオブジェクトほど高順位に出力する。

5. クロスメディア適合性フィードバックの例

本節では Flickr を用いた提案モデルの適用例を示す。

本来検索モデルとしては、各特徴量に関してオブジェクトを検索できるのが望ましいが、Flickr で可能な検索はテキスト検索のみなため、タグ集合のテキストクエリで検索をし、画像特徴量のクエリはランキングのみに用いる。

外観は思い浮かぶが、犬種名が思い出せない犬の画像を検索したい場合、図 3 のように、ただ単に”dog”というクエリで検索することになってしまう。

出力結果としてはオブジェクト、画像特徴量、タグ集合、画像特徴量の次元である色、タグ集合の次元であるキーワードなどがあり、ユーザはこれらの中から自由に選択し、正例や負例として与えることができる。

例えば、図 3 のように、2 つ目のオブジェクトの画像特徴量を正例として選んだとする。このとき、まず通常の適合性フィードバックとして、選択した画像特徴量を画像クエリに加える。それと同時に進行テキストクエリへの伝播は、選択した画像特徴量と一致するオブジェクトのテキスト特徴量を強めに、類似する画像特徴量を持つオブジェクトのテキスト特徴量を弱めに加える。

以上の操作により、2 つのクエリを修正することができたので、修正されたクエリで再検索、ランキング計算を行い、新たな検索結果を表示する。

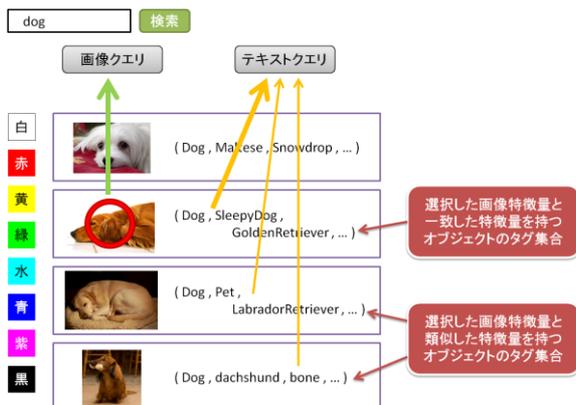


図 3 クロスメディア適合性フィードバックの適用例

6. まとめと今後の課題

本稿では画像特徴量やテキスト情報など、複数の特徴量をもつオブジェクトの検索において、適合性フィードバックを行う際に、従来の同一特徴空間内でのフィードバックだけでなく、画像からテキスト、テキストから画像のように特徴空間を跨いでフィードバックを行う、クロスメディア適合性フィードバックを提案した。また、オブジェクトレベルでの適合性フィードバックだけでなく、ベクトルレベル、次元レベルと正誤判定の対象を多様にし、柔軟な適合性フィードバックを行う手法を提案した。

次に、Flickr において画像検索を行う場合を考え、画像特徴量 CCV と画像に付与されたタグ集合を用い、クロスメディア適合性フィードバックの具体的な手法とともに、提案モデルの適用例を示した。

今後の課題としては、まず第一に実験評価を行ない、提案モデルの有効性を検証することが挙げられる。また、Flickr 以外で、例えば通常の Web ページを検索する際にも、ページのコンテンツを利用したクロスメディア適合性フィードバックが適用できないかといったように、提案モデルの汎用性を示すことも今後の課題である。

謝辞 本研究の一部は、京都大学 GCOE 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）によるものです。ここに記して謝意を表します。

文献

- [1] J. Rocchio: "Relevance Feedback in Information Retrieval", The SMART retrieval system: experiments in automatic document processing, pp. 313-323 (1971).
- [2] <http://www.flickr.com/>
- [3] G. Pass, R. Zabih and J. Miller: "Comparing Images Using Color Coherence Vectors", Proceedings of the 4th ACM international conference on Multimedia, pp. 65-73 (1997).
- [4] G. Salton and C. Buckley: "Improving retrieval performance by relevance feedback", Readings in information retrieval, pp. 355-364 (1997).
- [5] Y. Rui, T. Huang, M. Ortega and S. Mehrotra: "Relevance feedback: A Power Tool for Interactive Content-Based Image Retrieval", Proceedings of the IEEE Transactions on circuits and systems for video technology, 8, 5, pp. 644-655 (1998).