

研究速報

ユーザフィードバックを利用した逐次的学習による 検索意図に即した画像検索

中山 大輔[†] 保坂 忠明[†] (正員)

浜本 隆之[†] (正員)

Intent-Based Image Retrieval by Feedback-Assisted Sequential Learning

Daisuke NAKAYAMA[†], Nonmember, Tadaaki HOSAKA[†],
and Takayuki HAMAMOTO[†], Members

[†] 東京理科大学大学院工学研究科, 東京都
Graduate School of Engineering, Tokyo University of Science,
Tokyo, 162-8601 Japan

あらまし ユーザの検索意図に即した画像検索を実現する枠組みを提案する。ユーザフィードバックを利用することで、検索意図に即した学習データを逐次的に収集し、検索器を生成する。実際に画像検索手法を適用し、主観評価実験を通して提案枠組みの有効性を確認した。

キーワード 画像内容検索, 機械学習, 適合性フィードバック, 主観評価実験

1. ま え が き

画像内容に基づく検索では、人間の感覚とコンピュータによる認識の差であるセマンティックギャップを埋めることを目指し、機械学習が頻繁に用いられている [1]。それらの研究の多くは、画像特徴と画像内容 (クラス) からなる学習データをもとに、検索に有用となる認識規則や知識表現の獲得を目的としている。

機械学習をもとにして、ユーザにとって満足 of いく検索結果を得るためには、画像特徴量の選択以外に

(i) 画像間の距離測度が個々のユーザの検索意図を反映している

(ii) 用意される学習データ (特にクラスの割り当て) が個々のユーザの検索意図に即している

ことが重要となる。これまで研究されてきた適合性フィードバック [2] による手法の多くは、検索意図に対する検索結果の適合性をユーザフィードバックにより取得し、特徴量の重み付けの変更 [3] や機械学習における判別超平面の変更 [4] を一時的に行っている。これらの手法は、上記 (i) に焦点を当てているとみなすことができる。

本研究では、適合性フィードバックを上記 (ii) の実現のために利用することを考える。提案する画像検索の枠組みでは、ユーザの望むクラスに関するフィードバックを利用して、ユーザごとの検索意図に即した学

習データを徐々に蓄積し、再学習を行う。この再学習により、ユーザの検索意図が距離測度に反映されることになるため、提案枠組みは上記 (i) の要素も含んでいるといえる。別の視点から見れば、提案枠組みは、フィードバックを利用して逐次的に距離計量学習 [5] を行っているとみなすこともできる。

多くの従来研究のように、入手可能な画像データベース内の各画像にあらかじめ付与されたクラスに従って適合率や再現率を評価することは、本研究の趣旨にそぐわない。そこで、提案枠組みに従って生成される検索器がユーザの検索意図をどの程度、反映しているのかを主観評価実験により定量的に検証する。

2. フィードバックを利用した逐次的学習

提案する画像検索の枠組みを図 1 に示す。機械学習を用いた検索手法の多くが、提案枠組みへ適用可能であると考えられる。ただし、検索手法としては、

- ・クラス学習・識別を基礎としている
- ・画像間の距離を定義できる

ことを前提としている。以下に画像検索の手順を示す。

1. ユーザはクエリ画像を入力する。
2. 検索器は、クエリ画像との距離を算出し、初期検索画像をランキング順にユーザに提示する。
3. ユーザは、初期検索結果に満足すれば、そのまま検索を終了する。満足でない場合は、クエリ画像のクラスを既学習クラスの中から指定するか、新規クラスとして作成する。また、初期検索画像の中に自身の検索意図に合致した画像があれば、ユーザは適合画像

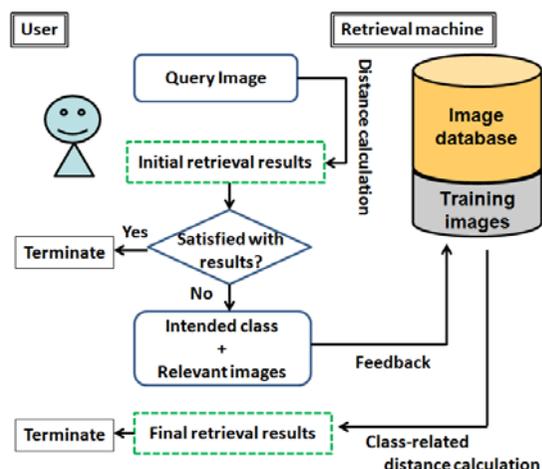


図 1 画像検索の枠組み
Fig. 1 Framework of the image retrieval.

としてそれらを選択する。

4. 検索器は、クエリと適合画像を指定クラスに対する学習サンプルとして獲得し、再学習を行う。

5. 検索器は、指定されたクラスを考慮に入れた距離を算出し、最終検索画像をランキング順にユーザに提示する。

以上の検索試行が繰り返されると、ユーザの検索意図に合致した学習データが徐々に蓄積されることになる。それに伴い検索器が発展していくものと期待し、次章で検証を行う。

3. 提案枠組みに対する有効性の検証

ここでは、先行研究 [6] の検索手法を提案枠組みへ適用し、4 人の被験者に対して主観評価実験を行う。なお、この検索手法は、画像の形状及び色特徴として SIFT の Bag of Features 表現 [7] と Color Local Auto Correlation [8] を利用し、学習器として正準相関分析を用いている。

3.1 検索器の概要

画像内容を表現するクラスベクトル（各次元がクラスに対応し、クラス概念が画像内に含まれているか否かを 1/0 により表現する）を \mathbf{x} (l 次元) で表し、画像特徴量ベクトルを \mathbf{y} (m 次元) で表現する。学習データに対し、

$$\mathbf{s} = A^T \mathbf{x} \quad (1)$$

$$\mathbf{t} = B^T \mathbf{y} \quad (2)$$

という二つの線形変換により得られるベクトル \mathbf{s}, \mathbf{t} (共に k 次元; $k = \min(l, m)$) の相関行列 Λ ($k \times k$ 次元) が対角行列になるという条件のもとで、そのトレースが最大となるように変換行列 A ($l \times k$ 次元), B ($m \times k$ 次元) を求める。これは、正準相関分析により実行することが可能である。更に、正準変量 \mathbf{t} から \mathbf{s} への線形回帰式は、

$$\hat{\mathbf{s}} = \Lambda \mathbf{t} \quad (3)$$

と求められる。ここで、記号 $\hat{\cdot}$ は、推定値であることを表している。 $l \leq m$ が成立しているとき、以上の関係から学習データ以外の画像に対するクラスベクトルは、

$$\hat{\mathbf{x}} = (A^T)^{-1} \hat{\mathbf{s}} = (A^T)^{-1} \Lambda \mathbf{t} \quad (4)$$

と推定される。ただし、ベクトル $\hat{\mathbf{x}}$ の各次元は 1/0 の 2 値ではなく実数となることに注意する。



図 2 データベースに含まれる画像例
Fig. 2 Image examples in our database.

提案枠組みのステップ 2 における画像間の距離 d は、クエリ画像及びデータベース内画像の推定クラスベクトル $\hat{\mathbf{x}}_q, \hat{\mathbf{x}}_t$ を用いて

$$d = \|\hat{\mathbf{x}}_q - \hat{\mathbf{x}}_t\|_1 \quad (5)$$

により定義される。ここで、 $\|\cdot\|_1$ は $L1$ ノルムを表す。更にステップ 5 では、被験者により指定された正解クラスの寄与を大きくし画像間の距離を再定義する。様々な方法が考えられるが、ここでは、ベクトル $\hat{\mathbf{x}}_t$ の正解クラスに対応する次元の値を α ($10^{-6} \leq \alpha \leq 1$; 区間外の値は近い方の境界値で近似する) で表し、

$$d' = \frac{\|\hat{\mathbf{x}}_q - \hat{\mathbf{x}}_t\|_1}{\alpha} \quad (6)$$

とした。指定された正解クラスの概念が検索対象の画像内に含まれている場合に α の値が大きくなることが期待でき、距離 d' の値は小さくなる。

ステップ 3 におけるクラスの指定では、登録済みのクラスラベルの一覧が表示され、被験者はその中から自身の感覚に合致するものを選択するか、若しくは新規のクラスラベルを登録することになる。

3.2 被験者による検索試行

検索対象となる画像データベースには、インターネット上で著作権フリーの画像素材として公開されている多様な画像 6100 枚を独自に収集し使用した。多くの従来研究や文献 [6] で用いられた Caltech-101 などでは、多くの画像のクラスが自明であり、ユーザ間で検索意図に差異が現れにくいと考えた。一方で、収集した画像群にはクラスがあらかじめ付与されていないが、図 2 に示した例のとおり、クラスの定義に個人差が生じやすい多数の画像が含まれている。

被験者に課す検索試行の条件を以下に示す。

- ・被験者によらない共通の 5 クラス各 5 枚の学習データを用いて検索器は初期化されている。

- ・被験者は、前節の手順を繰り返して計 100 回の検

表 1 主観評価実験 1 の結果 (平均評価値)
Table 1 Results of experiment 1 (Average scores).

| | 使用する画像検索器 | | | |
|-------|-----------|-------|-------|-------|
| | 検索器 A | 検索器 B | 検索器 C | 検索器 D |
| 被験者 A | 3.9 | 3.2 | 3.4 | 3.4 |
| 被験者 B | 3.4 | 3.8 | 3.4 | 3.0 |
| 被験者 C | 2.9 | 3.1 | 3.8 | 3.2 |
| 被験者 D | 2.6 | 2.3 | 2.5 | 3.8 |

索を行う。なお、各試行では、それまでの検索履歴や学習経過と無相関にランダムサンプリングされた 40 枚の画像の中から被験者はクエリ画像として 1 枚を選択する。

- ・被験者は、初期検索画像で満足することなく、クラスの指定・作成及び適合画像の選択まで行う。
- 初期検索画像は、上位 80 枚が提示される。
- 最大 3 枚まで適合画像を選択可能とする。

3.3 主観評価実験及び実験結果

全被験者の 100 回の検索試行が終了した後に、以下の二つの主観評価実験を被験者ごとに行った。

<主観評価実験 1 >

個々のユーザの検索意図に即した検索器が生成されているかどうかを検証する。手順を以下に示す。

1. 被験者は、ランダムサンプリングされた 40 枚の画像の中から新たにクエリを選択する。
 2. 被験者 4 名に対応する検索器から得られる初期検索画像 (各上位 30 枚) が被験者に提示される。この際、各結果がどの被験者の検索器から得られたものかは知らされない。
 3. 被験者は、四つの検索結果の各々に対して、5 段階の評価 (5:非常に良い, 4:良い, 3:ふつう, 2:悪い, 1:非常に悪い) を付ける。
- 各被験者は、1~3 の手順を 10 回行う。

被験者と検索器の組ごとに求めた平均評価値を表 1 に示す。全ての被験者が自らの検索器に対し最も高い評価値を与えていることが分かる。また、その値も 3.8~3.9 であり、クエリのクラスを考慮していない初期検索画像としては高い満足度であると考えられる。ある被験者の検索器から得られた初期検索画像の例を図 3 に示す。逐次学習の過程で、この被験者の定義したクラスの一つが「猫」という概念であったことを実験後に確認しており、その意図が検索結果に良好に反映されている。

<主観評価実験 2 >

検索を繰り返すことで検索精度が徐々に向上してい



図 3 検索結果例 1 (初期検索画像, 評価値: 4)

Fig. 3 Examples of retrieval results 1 (Score: 4).
※左上がクエリ画像, 上位検索結果 30 件を表示 (上段左より 1 位~6 位, 第 2 段左より 7 位~12 位, 以降同様).

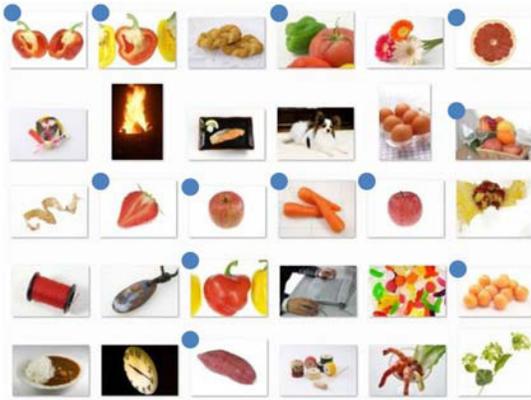
表 2 主観評価実験 2 の結果 (平均評価値)
Table 2 Results of experiment 2 (Average scores).

| | 50 回目終了時 | | 100 回目終了時 | |
|-------|----------|------|-----------|------|
| | 初期検索 | 最終検索 | 初期検索 | 最終検索 |
| 被験者 A | 3.8 | 3.8 | 4.0 | 4.2 |
| 被験者 B | 2.6 | 3.7 | 3.4 | 3.9 |
| 被験者 C | 2.9 | 3.8 | 3.5 | 4.0 |
| 被験者 D | 2.1 | 2.9 | 3.2 | 4.1 |

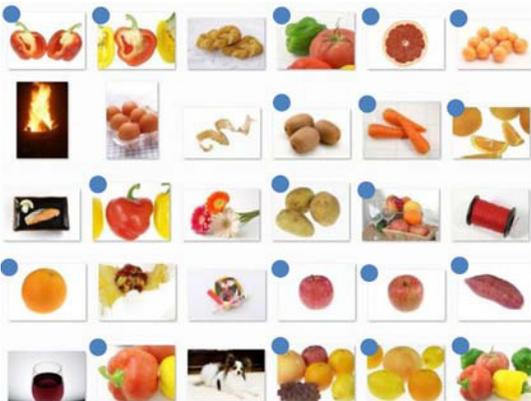
るかどうかを検証する。手順を以下に示す。

1. 被験者は、ランダムサンプリングされた 40 枚の画像の中から新たにクエリを選択する。
 2. 該当被験者の 50 回目と 100 回目の検索終了時における検索器の初期検索画像と最終検索画像 (各上位 30 枚) が被験者に提示される。この際、各結果がどの検索器から得られたものかは知らされない。
 3. 被験者は、四つの検索結果の各々に対して 5 段階の評価 (5:非常に良い, 4:良い, 3:ふつう, 2:悪い, 1:非常に悪い) を付ける。
- 各被験者は、1~3 の手順を 10 回行う。

検索器ごとに求めた平均評価値を表 2 に示す。全ての被験者が、初期検索画像と最終検索画像ともに 50 回目より 100 回目の方に高い評価値を与えていることが分かる。特に 100 回目の最終検索画像の評価値は 3.9~4.2 を示しており、高い満足度であると考えられる。100 回の検索終了後に評価値 5 が付けられた最終検索画像例を対応する初期検索画像とともに図 4 に示す。なお、ユーザの意図 (この被験者は「果物・野菜」の概念に対応するクラスを作成していた) に合致すると判断される画像の左上に丸印を付した。ユーザがクラスを指定することで、最終検索結果では、相対的によ



(a) 初期検索結果 (評価値: 3)
(a) initial retrieval results (score: 3)



(b) 最終検索結果 (評価値: 5)
(b) final retrieval results (score: 5)

図4 検索結果例2

Fig. 4 Examples of retrieval results 2.

※左上がクエリ画像，上位検索結果 30 件を表示 (上段左より 1 位～6 位，第 2 段左より 7 位～12 位，以降同様)。

り多くの望ましい画像が類似画像として上位に得られている。

上で述べた二つの主観評価実験より，ユーザフィードバックを利用することで検索意図を反映した学習データが収集され，検索性能の向上につながる事が確認された。

提案枠組みは，ユーザフィードバックとして適合画像の選択とクラスラベルの指定を課すため，前者のみを課す適合性フィードバックの手法と比較した場合にユーザの負担は大きいといえる．特に本研究では，主観評価実験のために，検索試行ごとに必ずフィー

ドバックを課していた．しかし，現実の運用の場面では，(i) ユーザが初期検索結果に満足しない場合にのみフィードバックを課す，(ii) クラスラベルの指定を省略し推定されたクラスベクトルで代用する，(iii) クラスラベルの指定のみを行わせる，ことなどによりユーザの負担を減じることも可能である．

4. むすび

本論文では，ユーザの検索意図に即した画像検索を実現する枠組みを提案した．提案枠組みでは，ユーザのフィードバックを用いて検索意図に即した学習データを蓄積し，徐々に検索精度を向上させていく．被験者に対する主観評価実験を通して，提案枠組みに従い効果的にユーザの検索意図を獲得できることを確認した．機械学習に基づく様々な検索手法の提案枠組みへの適用や主観評価によらない有効性の検証が今後の課題となる．

文 献

- [1] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognit.*, vol.40 no.1, pp.262-282, 2007.
- [2] X.S. Zhu and T.S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia System*, vol.8, no.6, pp.536-544, 2003.
- [3] Y. Rui and T. Huang, "Optimizing learning in image retrieval," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp.236-243, 2000.
- [4] L. Zhang, F. Liu, and B. Zhang, "Support vector machine learning for image retrieval," *Proc. IEEE Conf. on Image Processing*, pp.7-10, 2001.
- [5] T. Hertz, A. Bar-Hillel, and D. Weinshall, "Learning distance functions for image retrieval," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp.II-570-II577, 2004.
- [6] D. Nakayama, T. Hosaka, and T. Hamamoto, "Content-based image retrieval using SIFT and CILAC features," *Proc. Workshop on Picture Coding and Image Processing*, pp.103-104, 2010.
- [7] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bag of keypoints," *Workshop on Statistical Learning in Computer Vision, Proc. European Conference on Computer Vision*, pp.1-22, 2004.
- [8] T. Kobayashi and N. Otsu, "Color image feature extraction using color index local auto-correlations," *Proc. International Conference on Acoustics, Speech and Signal Processing*, pp.1057-1060, 2009.

(平成 23 年 2 月 28 日受付，7 月 2 日再受付)