

コミュニケーションを通じた物体を 特定するための特徴学習の構築

鈴木 隆二[†] 高橋 淳二[†] 戸辺 義人[†]

青山学院大学理工学部情報テクノロジー学科[‡]

1. はじめに

組み立て工場、医療現場、介護サービス、保育園などの現場における労働力不足に対してサービスロボットの代替が期待されている。しかし、現在のサービスロボットは、研究のデモンストレーションや人の関心を集めるといったことしか実現できておらず、実用化には程遠い。物体・環境認識、会話の理解、といった能力が不足しているためである。これらの個々の能力を個別に向上させることと同様に、全ての能力を統合して向上させることも重要である。本研究ではロボットの物体認識の学習とコミュニケーション方法の学習を両立して、インタラクティブな学習を行う LTIP(Learning Through Interaction with Person)を提案する。LTIPは、ユーザとロボットが協力することで実世界の中で物体探索を行う状況を想定する。ロボットは自分自身にとって未知な物を探索して欲しいという依頼に対して、Webを利用した候補画像の収集と人とのインタラクティブな絞り込みを行う。本稿では、LTIPの設計と評価実験結果について述べる。

2. 関連研究

Web画像を事前に自動収集し、一般物体認識のための学習画像データとして利用し、クラス分類を行うことを提案している研究¹⁾がある。本研究では、事前に画像を収集し、データベースの作成を行うのではなく、ユーザによって収集画像のキーワードを決定する。また、クラス分類を行うのではなく、画像のクラスタリングを行い、クラスタリングの結果に対するユーザからの情報に対して、再度反復してクラスタリングを行うことで、物体の外観を特定することを目指す。

3. LTIPの設計・実装

LTIPの主要な要素は(1)候補画像の収集、(2)インタラクティブによる絞り込みであり、以下

Learning Features through Communication in the Context of Finding an Object

[†] Ryuji SUZUKI, Junji TAKAHASHI, Yoshito TOBE

[‡] Aoyama Gakuin University

に各要素の内容を示す。

(1) 候補画像の収集

ロボットは、物を探して欲しいという要求に対して探してほしい物(以下「物」と表記)についての情報は持たない。そのため「物」のキーワードを元に、Bing Search APIを用いてWebから候補画像の収集を行う。収集した候補画像は画像サイズが均一でないので、縦横比を維持した状態で候補画像サイズの変更を行う。

(2) インタラクティブによる絞り込み

収集した候補画像は様々であり、絞り込みを効率よく行わないと、時間を要してしまう。そこで、①画像特徴抽出、②SOMを用いた分類、③選択による絞り込みという手順を踏むことで効率よく絞り込みを行う。以下でそれぞれについて述べる。

① 画像特徴抽出

収集した候補画像を区別するために、画像の特徴を用いる。用いた画像の特徴としては、以下の5つで、それぞれ色または形に関する特徴である。

● 色に関する特徴

- RGB値(3次元): 各画像のピクセルごとのRGB値の値を元にヒストグラムを作成し、R,G,B値それぞれの平均値を求める。
- 二値化(3次元): 画像の輝度を求め、閾値を超える輝度であるピクセルの割合を求める。3つの閾値を用いることで、次元数を3とした。

● 形に関する特徴

- Hog特徴(1次元): 光度の勾配方向・強度の算出を行い、ヒストグラムを求める。
- Hough変換(4次元): 画像内に含まれる円と直線の検出を行い、円と直線の数と長さを求める。
- Conner特徴(1次元): 画像内に含まれる物体のエッジを検出し、エッジの個数を求める。

② SOMを用いたクラスタリング

クラスタリングの方法として、SOM(Self-

Organize Maps)を用いる。SOMは教師なしのニューラルネットワークの一種で、入力層と出力層から成る。SOMは、相互に類似したものを自動でクラスタリングすることが可能で、特に視覚的に見た際にその傾向が表れる。入力層として、抽出した画像特徴を用いる。

③ 選択による絞り込み

分類した候補画像の中から各々のクラスを中心に一番近い画像を代表画像として提示する。提示した代表画像の中から「物」に一番近いものをユーザに選択してもらうことで絞り込みを行い、再度クラスタリングと選択を繰り返すことで「物」の画像を特定する。

LTIPの手順を以下に示す。

1. ロボットは「物」のキーワードを受け取る。
2. キーワードを元にAPIを用いて候補画像を収集する。
3. 候補画像のサイズを統一する。
4. 候補画像の特徴を抽出する。
5. 特徴ベクトルを元にSOMでクラスタリングする。
6. 各クラスを中心の画像を代表画像とし提示する。
7. ユーザは提示された画像の中で「物」に最も近い代表画像を選択する。
8. 選択画像が「物」であれば、絞り込みを終了。「物」でなければ、5に戻る

4. 評価実験

作成したシステムの評価を行うために、“cup”をキーワードとし、APIを用いて候補画像(図1)を収集した。収集した候補画像を用いて評価実験を行った。図1のうち一つを頭に思い浮かべた状態で、思い浮かべたものを特定するということを3つのTargetについて行った。Targetは、青、赤、緑色の四角で囲んだもので、それぞれT-1, T-2, T-3とする。

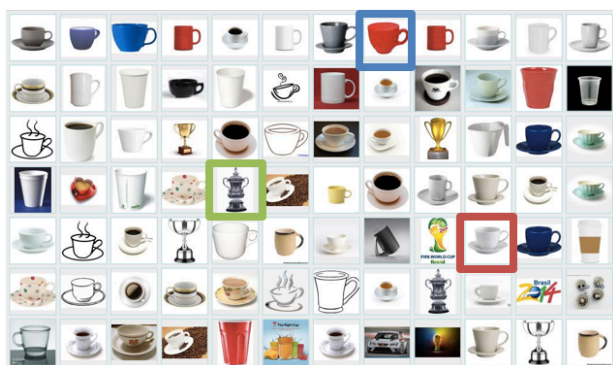


図1 実験に用いた「カップ」画像

実験は、色に関する6次元の特徴を用いた場合、色・形に関する12次元全ての特徴を用いた場合の2つの場合で、3つのTargetについて30回ずつ行った。その際に代表画像の選択回数が何回必要だったかをカウントし、平均値を求めた(表2)。また、6次元と12次元の場合のクラス分散比を $J(= \text{クラス外分散} / \text{クラス内分散})$ とし、 J の平均値を求めた(表1)。 J の値が高い程統計的にはよい分類が行えていることになる。

表1 クラス分散比(30回平均)

| | | |
|----------|-------------|-------------|
| | 6次元 | 12次元 |
| J (分散) | 26.2(127.6) | 25.0(119.9) |

表2 Target特定までの選択回数(30回平均)

| | | | |
|----------|----------|----------|----------|
| | T-1 | T-2 | T-3 |
| 6次元(分散) | 2.8(2.0) | 5.6(1.0) | 3.3(0.6) |
| 12次元(分散) | 3.3(2.1) | 4.4(0.7) | 1.3(0.4) |

実験結果のクラス分散比を表1に示す。6次元の場合と12次元の場合では違いがあまり見られなかった。

実験結果のTarget特定までの選択回数を表2に示す。次元数を6次元から12次元に変更すると、特定のための試行回数がT-1については減少し、T-2, T-3については増加した。これは、T-1は色が顕著であるが、形は顕著でないため、色・形に関する12次元の特徴を用いた場合では、T-1が際立たないからである。それに対して、T-3は色が顕著な上、形も顕著であるため、色・形に関する12次元の特徴を用いた場合では、T-3が際立ったものとなる。

以上を踏まえると、特定したい対象物や候補画像群によって用いる特徴を変更するべきであると考えられる。

5. むすび

本研究では、未知物体の外観を特定するためのシステムLTIPを提案し、評価実験を行った。特徴の次元数を変更しても、統計的には同等な分類性能であったが、LTIPで利用すると大きな違いがでることがわかった。

今後の課題としては、収集した候補画像から分類に用いるよりよい特徴の選択、ユーザが重要視する特徴の学習を行い、物体特定の簡易化を目指す。

参考文献

[1] 柳井啓司: 確率トピックモデルによるWeb画像の分類: The Japanese Society for Artificial Intelligence 2008