2R-06

画像に対する潜在的意味単語表現を用いた「かわいい」画像の分類方式

古宮 大暉 †　　秋吉 政徳 ‡
神奈川大学 †　　神奈川大学 ‡

1　はじめに

現在、機械学習を用いた画像の分類は広く研究されており、複雑なオブジェクトに対しても高い精度での分類を可能にしている。しかし、動物や人の顔などの具体的なオブジェクトを分類する技術が盛んに研究されている一方、「かわいい」などの人間の感覚が基準となる『感性語』によって表現される画像に対する分類は、研究があまりなされていない。

そこで本研究では、感性語の中でも特に表現が多様な「かわいい」という言葉によって表現される画像の分類方式を提案する。

2　提案方式

2.1　アプローチ

「かわいい」画像の分類方式を提案した研究がある [1]。この研究では、画像の形状や色の特徴をフィルターを用いて抽出し、定量的に特徴を表現した上で複数の機械学習分類器による比較実験を行い、その結果から「かわいい」画像に適した分類方式が提案されている。「かわいい」画像として 5 クラスの分類に対して、Random Forest を用いた実験にて最も高い精度 (70.2%) であった。

しかし、[1] では画像を構成する画素的な情報は扱っていたが、画像の示す内容を意味的に捉えることは行なっていなかった。画像の認識において内容理解は非常に重要な情報である。そこで本研究では、画像の意味的な構成内容を元とした分類方式を提案する。

Classification method for “kawaii” images using semantic words embedded in images
† Daiki Komiya, Kanagawa University
‡ Masanori Akiyoshi, Kanagawa University

2.2　構成

方式の構成を図 1 に示す。言語データベースから選定した単語によって構成される分類クラスを用いて、学習済み CLIP(Contrastive Language-Image Pre-training) による画像分類を行い、その際に獲得できる確率分布を元に、その画像に対する特徴ベクトルを抽出する。それらを入力に、複数の機械学習分類器による比較実験を行う。

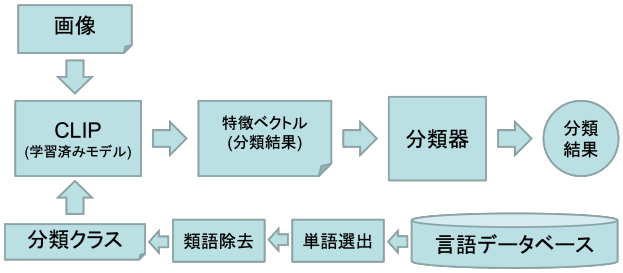

図 1　提案方式の構成

2.3　特徴ベクトルの作成

CLIP は画像と言語を同時処理し、zero-shot 学習で未学習クラスの画像を推測できる深層学習モデルである。学習済みの CLIP に分類クラスを与えることで、学習の有無に関わらずその分類が可能である。これにより、大規模データセットと学習のコストを要さずに画像の構成要素 (以降、オブジェクト) を認識するモデルを作成できる。本研究では、日本語学習済みモデル [2] を用いる。

CLIP は入力画像に対して、各クラスである確率分布を算出する。この値を、その画像が各クラスの要素を含む割合を示すものとし、画像内のオブジェクトを定量的に示す特徴ベクトルとする。

2.4　分類クラスの作成

日本語言語データベース [3] の単語を用いて分類クラスを作成する。本研究では、画像の特徴を

表現しやすい名詞に限定し、また頻繁に使用される単語はオブジェクトの認識に優れていると考え、使用頻度順に単語を選定した。選定した単語に類語除去を行ったのち分類クラスとする。

3　実験・結果

実験に用いた「かわいい」ジャンルと画像データセット数を以下の表 1 にまとめる。

表 1　**ジャンルとデータセット**

ジャンル名	データ数	ジャンル名	データ数
きもかわ	570	ゆるかわ	830
ぶさかわ	464	病みかわ	559
ゆめかわ	702		

実験には NN(Neural Network)、Random Forest 、Ada Boost 、SVM(Support Vector Machine) の 4 種の分類器を用いる。また、5 分割交差検定を採用し結果を比較した。

3.1　本実験

各分類器機、分類クラスを 100 単語から 100 刻みに 1000 単語までを用いた際の比較実験を行う。

実験結果を図 2 にまとめる。NN を用いた分類クラス 500 単語にて、全体で最も高い精度 (71.2%) となった。この結果は、従来研究 [1] の精度 (70.2%) を僅かながら上回った。

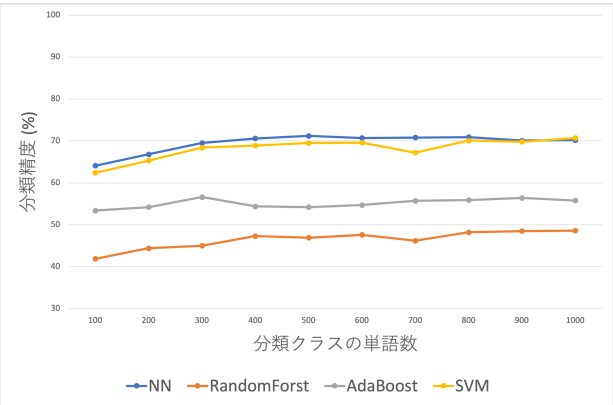

図 2　**実験結果**

3.2　追加実験

単語選定の効果を図るため、無作為に選定した名詞に対して同様の実験を行った。結果として、NN を用いた分類クラス 700 単語の場合に、全体で最も高い精度 (66.5%) となった。

4　考察

図 2 の実験結果から、特徴量として NN と SVM に本分類では優位性があることが確認できた。また、高い精度で分類できた NN、SVM について、一定の単語数以降に精度の向上が見られなかった。これは使用頻度上位の単語のみで十分に分類ができていたためと考えられ、単語選定が有効であったことが示されたと考えられる。また、単語選定の有効性は追加実験の結果からも確認できた。これは、人間が既知のオブジェクトの組み合わせから未知のオブジェクトを理解する方法に近い形で画像を認知できたとも考える。

5　おわりに

実験結果から、提案方式は従来研究 [1] 以上の精度で「かわいい」画像の構成内容を元に分類が可能であることを確認できた。特に、データベースからの単語選定の効果が有効に働いたためと考えられる。今後は、従来研究 [1] との組み合わせや説明可能 AI への応用について検討したい。

参考文献

[1] 古宮大暉, 古渡翔太, 秋吉政徳, “異種の特徴量フィルタを用いた「かわいい」画像の分類方式”, 情報処理学会, 第 85 回全国大会, 2U-07 (2023).

[2] シーン 誠, 趙 天雨, 沢田 慶, “日本語における言語画像事前学習モデルの構築と公開”, MIRU2022, IS1-80(2022).

[3] 松下達彦 (2011)「日本語を読むための語彙データベース（VDRJ） Ver. 1.0 （研究用）」, http://www17408ui.sakura.ne.jp/tatsum/database/VDRJ_Ver1_1_Research_Top60894.xlsx, (ダウンロード日：2023 年 11 月 10 日).