

属性情報の不確実性を考慮した未観測カテゴリ分類に関する研究

情報数理応用研究

5219C014-0 阪井優太

指導教員 後藤正幸

Classification of Unseen Category Considering Attribute Uncertainty

Yuta Sakai

1. 研究背景・目的

機械学習における分類問題は、データの特徴量とカテゴリを表すラベルをペアとした学習データの集合を用いて分類器を学習し、カテゴリが未知のテストデータのラベルを推定する。しかし、一般的な分類器では学習データ集合に含まれないカテゴリを推定することはできない。そのため、学習データを収集する際、識別したい全てのカテゴリを含む学習データの取得が必要となるが、これらを用意できない場合もある。

このような問題に対しては、想定しうる全てのカテゴリに共通する補助情報を活用し、転移学習を行うことで解決できる場合もある。この場合の転移学習とは、学習データのカテゴリから得られた知識をテストデータのカテゴリ分類タスクへと活用することを意味する。このデータに共通する補助情報を活用することで、(実際のデータには出現すると想定されるが)学習データ内では未観測なカテゴリを推定可能とするゼロショット学習 (Zero-shot Learning) [1] がある。このゼロショット学習における補助情報には、特徴量とカテゴリの関係性を間接的に表現可能なものを用いられる。例えば、動物の画像を分類する問題では、白い、飛ぶ、肉食などの様々な特徴を補助情報として用いる。本研究では、この補助情報として属性情報を活用する属性ベースゼロショット学習を対象とする。

この属性ベースゼロショット学習の代表的な手法として、Direct Attribute Prediction (DAP) モデル [2] がある。DAP モデルでは、予め学習データ集合とテストデータ集合で想定しうる全てのカテゴリについて、属性との関係を定義する。そして特徴量から全ての属性の有無を確率的に推定し、その属性情報のもとで所属確率が高いテストカテゴリに所属すると推定する確率的識別モデルである。

一方、属性によって特徴量への現れやすさは異なると考えられる。例えば画像の特徴量には、色彩情報と関係する属性は表現されやすく、それ以外の属性は表現されにくいと考えられる。DAP モデルは、特徴量に現れやすい属性を適切に学習するが、特徴量に現れにくい属性を適切に学習することが難しく、テストデータに対してカテゴリの推定を行う際、推定精度を低下させてしまう場合がある。また DAP モデルは、各属性値を属性の有無を表す 2 値でしか扱えないため、カテゴリ間での各属性値の大小関係を考慮することができない。

そこで本研究では、特徴量から連続的な属性値を分布として推定する属性ベースゼロショット生成モデルを提案する。これにより、各テストデータの特徴量から推定した属性値の分布のうち、推定結果の不確実性が高い属性を発見し、その属性を考慮しないことで高い分類精度を実現する。最後にベンチマークデータを用いたシミュレーション実験により DAP モデルと比較することで、提案モデルの有効性を示す。

表 1: 属性とカテゴリの例

特徴量				
カテゴリ		polar bear	tiger	otter
属性	white	95.62	30.12	0.00
	big	85.49	78.36	11.25
	flvs	0.00	0.00	0.00
	smart	16.25	36.59	27.22

2. 準備

2.1. 問題設定

属性ベースゼロショット学習では、予め学習データとテストデータにおいて想定しうるカテゴリと属性情報の関係性を定義する。属性とカテゴリの関係性の例を、表 1 に示す。ここで用いる属性情報とは人間が各カテゴリに対して付与できる質的な特徴である。表 1 の例では、polar bear や tiger というカテゴリが与えられており、各カテゴリに対して white や big といった属性との関係性の強さを表す値を与える。

属性ベースゼロショット学習における問題設定を定義する。特徴量空間 $\mathcal{X} \subseteq \mathbb{R}^d$ 内の特徴量を $\mathbf{x} \in \mathcal{X}$ 、その特徴量 \mathbf{x} に対応するカテゴリラベルを $y \in \mathcal{Y}$ とし、 \mathcal{Y} は K 個の加算有限集合 $\mathcal{Y} = \{l_1, \dots, l_K\}$ であるとする。これらのカテゴリラベルは、学習データもしくはテストデータのどちらかで生起するものとする。ゼロショット学習の目的は、学習データの特徴量とカテゴリラベルの関係性を学習し、テストデータの特徴量からカテゴリラベルを推定することである。しかしながら学習データとテストデータのカテゴリ集合が異なる問題設定であるため、両カテゴリラベルに共通する属性情報を導入し、属性情報を介したカテゴリの推定を図る。属性情報を表す属性空間 $\mathcal{A} \subseteq \mathbb{R}^M$ 内の属性を $\mathbf{a} = (a_1, \dots, a_M) \in \mathcal{A}$ と定義し、 a_v は v 番目の属性を表す値とする。

いま、学習データ集合を $\mathcal{D}_s = \{\mathbf{x}_{sn}, y_{sn}, \mathbf{a}_{sn}\}_{n=1}^{N_s}$ 、テストデータの特徴ベクトル集合を $\mathcal{D}_t = \{\mathbf{x}_{tm}\}_{m=1}^{N_t}$ と表す。ただし、 $y_{sn} \in \mathcal{Y}_s$ を学習データのカテゴリ、 $y_{tm} \in \mathcal{Y}_t$ をテストデータのカテゴリとする。 $\mathcal{Y}_s, \mathcal{Y}_t$ はそれぞれ、 $\mathcal{Y}_s = \{l_1, \dots, l_J\}$ 、 $\mathcal{Y}_t = \{l_{J+1}, \dots, l_K\}$ とし、学習データとテストデータのカテゴリは重複しないものとする ($\mathcal{Y}_s \cap \mathcal{Y}_t = \phi, \mathcal{Y}_s \cup \mathcal{Y}_t = \mathcal{Y}$)。学習データの属性 \mathbf{a}_{sn} と推定されるテストデータの属性 \mathbf{a}_{tm} は、共に M 個の属性ラベルの有無を表す M 次元ベクトル $\mathbf{a}_{sn}, \mathbf{a}_{tm} \in \mathcal{A}$ で定義されるものとする。この属性は各カテゴリに対して一意に与えられるものとし、カテゴリラベル l_k に対して $\mathbf{a}^{l_k} = (a_1^{l_k}, \dots, a_M^{l_k})$ を事前に定義する。そのカテゴリラベルと属性の事前情報を活用することで、テストデータのカテゴリ推定を可能とする。

ゼロショット学習では、学習データ集合 $\mathcal{D}_s = \{\mathbf{x}_{sn}, y_{sn}, \mathbf{a}_{sn}\}_{n=1}^{N_s}$ を用いて、テストデータの特徴ベクトル集合 $\mathcal{D}_t = \{\mathbf{x}_{tm}\}_{m=1}^{N_t}$ の属性情報 $\{\mathbf{a}_{tm}\}_{m=1}^{N_t}$ を推定し、この情報からテストカテゴリ $\{y_{tm}\}_{m=1}^{N_t}$ を推定する。

2.2. 関連研究

属性ベースゼロショット学習のベースとなる研究としては、Lampert らが提案した DAP モデルが知られている [2]。この DAP モデルは、データの特徴量から属性情報の有無を個別の識別器によって学習し、その識別結果を用いて確率的に属性からカテゴリを推定する手法である。この手法は、各属性情報が 2 値で与えられることを仮定している。表 1 の例では、polar bear と tiger と otter を white の属性で比較した場合、属性値を 2 値で扱うために polar bear は 1, tiger と otter は 0 とし、その値の大小で 2 値化を行う。

一方、属性ベースゼロショット学習では、学習データ数を十分に収集できないこともしばしば想定されている。そのため学習データが少数の場合を対象とするメタ学習と問題設定が類似しており、その代表的なモデルとして、MAML や Neural Process [3] が知られている。しかし、分類対象であるテストデータの特徴ベクトルとカテゴリの関係性を他の学習データによって学習する必要がある点でアプローチが異なる。

また、属性情報は多様な情報が与えられる。そのためデータの特徴量と属性の関係性を学習する際、その特徴量から学習が容易な属性と困難な属性が混在する。例えば、画像特徴量からは視覚的な属性情報が現れやすいが、それ以外の属性情報は現れにくいと考えられる。この問題点に対する類似した着眼点の研究として Parikh ら [4] は、人間にとって理解しやすい属性かどうかを考慮したモデルを提案している。しかしこの方法では、画像特徴量に対する人間の主観的な属性の選択を予め行うため、画像以外の一般的な問題設定には適さない。

3. 提案モデル

3.1. 概要

本研究では、特徴量から連続的な属性値を分布として推定する属性ベースゼロショット生成モデルを構築する。属性値を分布で推定することで、不確実な推定結果の属性情報を考慮せずカテゴリの推定が可能となる。このモデルは、特徴量 \mathbf{x} から属性 \mathbf{a} の推定と、それを用いたカテゴリ y の推定の 2 段階で構成されている。

特徴量 \mathbf{x} から属性 \mathbf{a} を推定する際、データの特徴量に対する各属性の現れやすさは異なると考えられる。そのため、推定した属性 \mathbf{a} からカテゴリ y を推定する際にすべての属性 a_1, \dots, a_M を同様に用いると、特徴量から現れにくい属性の推定結果がカテゴリの推定精度を低下させてしまう場合がある。そこで、この属性 \mathbf{a} の現れやすさを確率分布 $p(\mathbf{a}|\mathbf{x})$ で表現し、この分布を推定する。この属性分布 $p(\mathbf{a}|\mathbf{x})$ を各属性 a_v ごとに平均 μ_v と分散 σ_v^2 を持つような多次元正規分布で表現することにより、各属性のもとで得られる特徴量の不確実性をばらつきによって表現することが可能となる。すなわち、推定した属性の分散 σ_v^2 が大きいときは、データの特徴量 \mathbf{x} から推定した属性 a_v の平均値 μ_v が属性値を正しく表現している可能性が低く、推定結果の不確実性が高いと考えられる。そこ

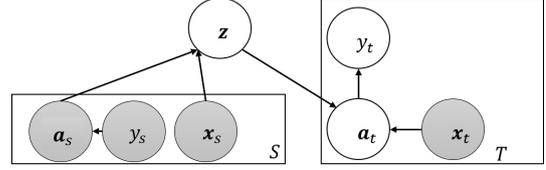


図 1: 提案モデルのグラフィカルモデル

で、推定した分布の分散が大きい属性は考慮せずにカテゴリの推定を行う方法を考える。

具体的には、次の 2 段階によりカテゴリ推定を行うモデルを提案する。まず、データの特徴量 \mathbf{x} から属性 \mathbf{a} を推定する際、各属性 a_v の平均 μ_v と分散 σ_v^2 を推定する。その後、推定した属性分布 $p(\mathbf{a}|\mathbf{x})$ を考慮することでカテゴリ y の推定を行う。データの特徴量 \mathbf{x} から属性 \mathbf{a} を推定する部分は、特徴量から属性を推定する関数 $f: \mathcal{X} \rightarrow \mathcal{A}$ を生成する確率過程のモデルをニューラルネットワークを用いて構築し、各属性 a_v において平均 μ_v と分散 σ_v^2 を持つ属性分布 $p(\mathbf{a}|\mathbf{x})$ を推定する。これにより、カテゴリの推定部分では各データの特徴量に対応した属性推定の不確実性を考慮し、テストデータのカテゴリ推定に有用な属性を絞り込む。提案モデルのグラフィカルモデルを図 1 に示す。なお、図 1 中の色付き変数は観測変数、 z は潜在変数、 (\mathbf{a}_t, y_t) は推定される変数を表す。また、 S, T はそれぞれのデータ数を表す。

3.2. 提案モデルの定式化

まず、以降のモデルの説明において、全データ N 個のデータ系列の特徴量を $\mathbf{x}_{1:N}$ 、属性を $\mathbf{a}_{1:N}$ と定義する。提案手法では、特徴量から属性への写像を行う関数 f の出力が正規分布に従って生成されると仮定し、テストデータ \mathbf{x}_t に対する属性分布 $\hat{p}(\mathbf{a}_t|\mathbf{x}_t)$ を推定する。その後、推定した属性の分布 $\hat{p}(\mathbf{a}_t|\mathbf{x}_t)$ からカテゴリ y_t の推定を行う。本節では、属性分布の推定を行う部分を定式化する。

個々のデータには観測ノイズが含まれるため、属性の出現度 \mathbf{a}_i は多次元正規分布 $\mathcal{N}(f(\mathbf{x}_i), \text{diag}[\sigma_i^2])$ に従うものとする。ここで、 $\text{diag}[\sigma_i^2]$ は対角要素が $\{\sigma_v^2\}_{v=1}^M$ の対角行列である。 $\mathbf{x}_{1:N}$ が与えられたとき、ある確率過程 $f: \mathcal{X} \rightarrow \mathcal{A}$ に従う属性の確率分布を式 (1) で定義する。

$$p(\mathbf{a}_{1:N}|f, \mathbf{x}_{1:N}) = \prod_{i=1}^N \mathcal{N}(\mathbf{a}_i|f(\mathbf{x}_i), \text{diag}[\sigma_i^2]). \quad (1)$$

また、データの特徴量 \mathbf{x} と属性情報 \mathbf{a} の生成過程を表現する潜在変数 z を導入し、その確率を式 (2) で定義する。

$$p(z, \mathbf{a}_{1:N}|\mathbf{x}_{1:N}) = p(z) \prod_{i=1}^N \mathcal{N}(\mathbf{a}_i|g(\mathbf{x}_i, z), \text{diag}[\sigma_i^2]). \quad (2)$$

ただし、 z は高次元空間を仮定した潜在変数であり、 z を固定したときの関数を $f(\mathbf{x}_i) = g(\mathbf{x}_i, z)$ とする。モデル上ではニューラルネットワーク $g(\cdot)$ を用いて、事前分布のパラメータ z と推定したい特徴量 \mathbf{x}_i のもとで、属性 \mathbf{a}_i を推定する関数とする。

属性の対数尤度の変分下限の導出結果を式 (3) で表す。

$$\begin{aligned} & \log p(\mathbf{a}_{1:N} | \mathbf{x}_{1:N}) \\ & \geq \mathbb{E}_{q(\mathbf{z} | \mathbf{x}_{1:N}, \mathbf{a}_{1:N})} \left[\sum_{i=1}^N \log p(\mathbf{a}_i | \mathbf{x}_i, \mathbf{z}) + \log \frac{p(\mathbf{z})}{q(\mathbf{z} | \mathbf{x}_{1:N}, \mathbf{a}_{1:N})} \right]. \end{aligned} \quad (3)$$

式 (3) 右辺の第 1 項は、潜在変数 \mathbf{z} のもとで得られる属性の推定値の対数尤度を示す。第 2 項は負の $q(\mathbf{z} | \mathbf{x}_{1:N}, \mathbf{a}_{1:N})$ と $p(\mathbf{z})$ の KL ダイバージェンスを示す。この変分下限 (ELBO) の最大化を行うことでパラメータを推定する。

3.3. モデルの学習

学習時には式 (3) の変分下限最大化を学習データで行うため、学習データ集合 \mathcal{D}_s を $\mathcal{D}_{tr} = \{\mathbf{x}_{1:j}, \mathbf{a}_{1:j}\}$ と $\mathcal{D}_{va} = \{\mathbf{x}_{j+1:N_s}, \mathbf{a}_{j+1:N_s}\}$ に分割して利用する。

\mathcal{D}_{tr} と \mathcal{D}_{va} を用いることで特徴量から属性への写像を行う関数 f を学習し、 \mathbf{a}_i の分布を推定する。

\mathcal{D}_{tr} と \mathcal{D}_{va} から推定する属性の周辺対数尤度の変分下限を式 (4) で表す。

$$\begin{aligned} & \log p(\mathbf{a}_{j+1:N_s} | \mathbf{x}_{1:N_s}, \mathbf{a}_{1:j}) \\ & \geq \mathbb{E}_{q(\mathbf{z} | \mathbf{x}_{1:N_s}, \mathbf{a}_{1:N_s})} \left[\sum_{i=j+1}^{N_s} \log p(\mathbf{a}_i | \mathbf{x}_i, \mathbf{z}) + \log \frac{p(\mathbf{z} | \mathbf{x}_{1:j}, \mathbf{a}_{1:j})}{q(\mathbf{z} | \mathbf{x}_{1:N_s}, \mathbf{a}_{1:N_s})} \right]. \end{aligned} \quad (4)$$

式 (4) では、式 (3) における $p(\mathbf{z})$ を $p(\mathbf{z} | \mathbf{x}_{1:j}, \mathbf{a}_{1:j})$ として表現する。この第 1 項において属性の推定値と真値の誤差を測り、第 2 項において負の $q(\mathbf{z} | \mathbf{x}_{1:n}, \mathbf{a}_{1:n})$ と $q(\mathbf{z} | \mathbf{x}_{1:j}, \mathbf{a}_{1:j})$ の KL ダイバージェンスを測る。このとき $q(\mathbf{z} | \mathbf{x}_{1:n}, \mathbf{a}_{1:n})$ は \mathcal{D}_{tr} と \mathcal{D}_{va} を合わせたデータのもとでの潜在変数の分布を表す。この推定誤差と KL ダイバージェンスを最小化するモデルを学習する。推定誤差の計算時は $q(\mathbf{z} | \mathbf{x}_{1:j}, \mathbf{a}_{1:j})$ から \mathbf{z} をサンプリングして計算する。モデルの学習の流れを以下に示す。

- Step1** \mathcal{D}_s から \mathcal{D}_{tr} と \mathcal{D}_{va} をサンプリング
- Step2** \mathcal{D}_{tr} のデータを入力し、式 (4) で学習したパラメータを用いて潜在変数の分布 $p(\mathbf{z} | \mathcal{D}_{tr})$ を推定
- Step3** $p(\mathbf{z} | \mathcal{D}_{tr})$ の正規分布から \mathbf{z}_{sample} を生成
- Step4** \mathbf{z}_{sample} と推定したいデータ $\mathbf{x}_{j+1:n}$ を入力して、属性 $\mathbf{a}_{j+1:n}$ を推定
- Step5** Step4 の推定結果を用いて、式 (4) を最大化するようにニューラルネットワークを更新

この Step1 から Step5 までを繰り返すことでモデルのパラメータを推定する。

3.4. 目的タスクの推定

カテゴリの推定は、テストデータの特徴量からの属性推定と、それを用いたカテゴリ推定の 2 ステップにより行われる。

3.4.1. 属性の推定

3.3 節で学習したモデルを用い、全学習データ $\mathcal{D}_s = \{\mathbf{x}_{sn}, \mathbf{y}_{sn}, \mathbf{a}_{sn}\}_{n=1}^{N_s}$ とテストデータ $\mathcal{D}_t = \{\mathbf{x}_{tm}\}_{m=1}^{N_t}$ からテストデータの属性 \mathbf{a}_{tm} を推定する。ここで推定される属性

\mathbf{a}_{tm} は、平均ベクトル $\boldsymbol{\mu}_{tm}$ と分散を表す対角行列 $diag[\boldsymbol{\sigma}_{tm}^2]$ の 2 つのパラメータを出力する。属性推定の流れを以下に示す。

Step1 \mathcal{D}_s のすべてのデータを入力して、3.3 節の Step2 で学習したニューラルネットワークにより $p(\mathbf{z} | \mathcal{D}_s)$ を計算

Step2 $p(\mathbf{z} | \mathcal{D}_s)$ の正規分布から平均値 $\boldsymbol{\mu}_z$ を算出

Step3 $\boldsymbol{\mu}_z$ と \mathbf{x}_{tm} を 3.3 節の Step4 で学習したニューラルネットワークに入力して、属性 \mathbf{a}_{tm} を推定

3.4.2. カテゴリの推定

属性からのカテゴリ推定には、3.4.1 項で推定した属性 \mathbf{a}_{tm} を用いる。この \mathbf{a}_{tm} をもとに、テストカテゴリに対応する属性情報ベクトルを用いて分類する。このときカテゴリの推定は、属性の推定精度によって大きく影響を受けると考えられ、属性の推定精度は属性と各データの特徴量との関係性の強さが影響すると考えられる。

そこで属性の推定値の分散が小さい、すなわち信頼度が高いと想定される属性を用いて推定を行う。具体的には、テストデータによって推定した属性分布 $\hat{p}(\mathbf{a}_{tm} | \mathbf{x}_{tm})$ の平均ベクトル $\boldsymbol{\mu}_{tm}$ と分散を表す対角行列 $diag[\boldsymbol{\sigma}_{tm}^2]$ を用いる。この対角行列 $diag[\boldsymbol{\sigma}_{tm}^2]$ の v 番目の属性の分散 σ_v^{tm2} が大きい場合、推定の不確実な属性と考えられる。

属性の空間内において、テストカテゴリ l_k に事前に定義した属性ベクトル \mathbf{a}^{l_k} と推定した属性の平均ベクトル $\boldsymbol{\mu}_{tm}$ の距離を式 (5) により計算し、カテゴリの推定を行う。

$$\hat{y}_{tm} = \arg \min_{l_k} \sum_{v=1}^{M_u} \frac{a_v^{l_k} - \mu_v^{tm}}{\sigma_v^{tm2}}. \quad (5)$$

M_u は、カテゴリ推定に用いる特徴量数である。 M_u の値が大きいほど考慮する属性の数は多く、小さいほど考慮する属性の数は少なくなる。考慮する属性の数 M_u は、テストカテゴリの数を考慮しつつ適切な値を選ぶ必要がある。

4. 既存研究との精度比較実験

提案手法の有効性を検証するためにゼロショット学習における分類結果を DAP と比較する。また、提案手法においてテストデータによって考慮される属性が異なるため、カテゴリ推定に使用される頻度の高い属性と使用される頻度の低い属性をそれぞれ上位 5 件ずつ確認し、分析を行う。

4.1. 実験データ概要

以下では、Animals with Attributes データセット [2] を使用して実験を行った。このデータは属性がカテゴリごとに付与されており、個々の画像と関係のない属性も含まれている。画像データ数 30,475、カテゴリ数 50、属性数 85 であり、カテゴリの種類は動物である。画像の特徴量には、vgg19 を用いて抽出した 4096 次元の画像特徴量を活用する。

4.2. 実験条件

学習タスクのカテゴリ数は 40、目標タスクのカテゴリ数は 10 とし、両タスクで共通するカテゴリを持たないものとする。学習に用いるデータ数を各カテゴリ 10,20,30,40,50 まで 10

件ずつ変化させ実験を行った。またテストデータは、各カテゴリ 10 件ずつランダムに 5 回抽出し、各推定カテゴリの平均正解率で評価を行う。従来手法としてニューラルネットワークを用いた DAP (NN-DAP) と比較する。提案手法のハイパーパラメータは、潜在表現 z の次元数として 1000、ニューラルネットワークはエンコーダ、デコーダ共に中間層は 3 層とし、それらのユニット数は全て 1000 とした。またカテゴリ推定に用いる属性の数は $M_u = 50$ とした。NN-DAP のハイパーパラメータは、ニューラルネットワークの中間層を 3 層とし、それらのユニット数も全て 1000 とした。ニューラルネットワークの学習には Adam を用い、学習率を 0.003 とした。

4.3. 実験結果

提案手法と NN-DAP のカテゴリの平均正解率を比較した結果を表 2 に示す。

表 2: 各カテゴリのデータ数に応じた平均正解率

	データ数				
	10	20	30	40	50
NN-DAP	0.284	0.302	0.326	0.356	0.384
提案手法	0.502	0.524	0.558	0.582	0.602

いずれのデータ数においても NN-DAP と比較して、提案手法の平均正解率が高い値を示し、提案手法では学習データ数が少ない場合においても高い精度を示した。

また、推定したテストカテゴリにおいてカテゴリ分類時に考慮されている頻度の高い属性と考慮されている頻度の低い属性を以下の表 3 に示す。

表 3: カテゴリ推定に使用される属性の頻度

順位	考慮している頻度が高い	考慮している頻度が低い
1	red	tail
2	flys	slow
3	froager	brown
4	bush	solitary
5	skimmer	black

カテゴリ推定に使用される頻度が高い属性、つまり推定結果の信頼度が高い属性として、色彩情報の red (赤色) や画像特徴量の背景などによって推定可能と考えられる flys (飛ぶ), bush (茂み), skimmer (水面を飛ぶ) などが得られた。逆にカテゴリ推定に使用される頻度が低い属性、つまり推定結果の信頼度が低い属性として、どのカテゴリにも共通して現れやすい色彩情報である brown (茶色), black (黒色) や、画像特徴量から得られにくい tail (尻尾), slow (遅い) などが得られた。

4.4. 考察

まず、表 2 のゼロショット学習を行った場合の平均正解率について考察を行う。表 2 に示されたように、提案手法の平均正解率は、NN-DAP に比べて大幅に改善されている。提案手法が NN-DAP と比較して平均正解率で勝れている理由として以下の 2 点が考えられる。

1 点目は、画像特徴量と属性の値の関係性を連続値で推定可能となったことが、精度向上に寄与したことである。

2 点目は、カテゴリ推定を行う際に曖昧な推定結果の属性を用いないことによる効果である。今回の実験においてカテゴリ推定に用いる属性の数 M_u を最初に 50 と決めて実験を行っていた。この属性の数 M_u は本来信頼できる属性の数を何らかの基準のもとで設定することが好ましい。そこで曖昧な推定結果の属性を棄却したことによる影響を確認するために、カテゴリ推定に用いる属性の数 M_u を変化させた場合の提案手法の平均正解率の変化についても実験をおこなった。 $M_u = 40, 50, 60, 70, 85$ として、各カテゴリ 10 件 (10-shot), 50 件 (50-shot) で学習した結果を表 4 に示す。

表 4: 属性数 M_u を変化させた平均正解率

	属性数 M_u				
	40	50	60	70	85
10-shot	0.464	0.502	0.484	0.480	0.472
50-shot	0.572	0.602	0.610	0.604	0.596

属性数として $M_u = 50, 60$ としたときに良い平均正解率を示した。特に学習データ数が少ない場合 (10-shot), カテゴリの推定時に考慮する属性数を少なめに選択することで平均正解率が向上している。その一方で、学習データ数がある程度得られている場合 (50-shot), 10-shot の場合と比較して考慮すべき属性数を少し増やした方が平均正解率が上がった。これは学習データが増加したことでうまく特徴を捉えられる属性が増えたと考えられ、少数のデータでは学習の難しい属性があると解釈できる。属性によって画像特徴量のパターンが多様なものと単調ものが混在していると考えられる。

5. まとめと今後の課題

本研究では、属性情報の推定の不確実性を考慮した属性ベースのゼロショット生成モデルを提案した。従来手法との比較実験を行うことでその有効性を示した。また、提案手法の属性情報の推定結果を考察することで、特徴量から推定しやすい属性情報についても考察を行った。

さらなる問題設定として、学習データも含めたカテゴリ分類である一般化ゼロショット学習への拡張が考えられる。

参考文献

- [1] Lampert, C. H., Nickisch, H. and Harmeling, S. "Learning to detect unseen object class by between-class attribute transfer," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.951–958, 2009.
- [2] Lampert, C. H., Nickisch, H. and Harmeling, S. "attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2014-03, Vol.36, No.3, pp.453–465, 2014.
- [3] Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S. M. and Teh, Y. W. "Neural processes," *ICML Workshop on Theoretical Foundations and Applications of Deep Generative Models*, 2018.
- [4] Parikh, D. and Grauman, K. "Interactively building a discriminative vocabulary of nameable attributes," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1681–1688, 2011.