

Prototypical Network の中間層における 2 クラス間共通特徴の学習

佐藤 亮太[†] 藤田 悟[†]法政大学大学院 情報科学研究科[†]

1. まえがき

近年、深層学習においてZero-shot Learningという学習手法の研究が進められている。Zero-shot Learningとは、事前学習した既知のデータの知識を利用して未知のデータを処理する能力を獲得する学習手法である。

先行研究のPrototypical Networkは、入力データを低次元の埋め込み空間に出力するEmbeddingを行い、埋め込み空間内で各クラスのクラスターが離れるように学習する。本稿では、Prototypical Networkを改良し、未知のクラスのクラスターをよりはっきりと分離することができる学習手法を提案する。

2. 関連研究

2.1. Prototypical Network

Prototypical Network[1]とは、データの特徴を低次元の埋め込み空間に出力し、各クラスのプロトタイプ(代表点)の座標と入力データの座標によって分類を行うニューラルネットワークである。クラス k のいくつかのサンプルを S_k とすると、プロトタイプ \mathbf{c}_k は式(1)に示すように、 S_k の平均で求める。

$$\mathbf{c}_k = \frac{1}{|S_k|} \sum_{(\mathbf{x}_i, y_i) \in S_k} f_\phi(\mathbf{x}_i) \quad (1)$$

このプロトタイプを利用し、入力データがどのクラスのプロトタイプに近いかによって分類を行う。入力データ \mathbf{x} がクラス k である確率 $p_\phi(y = k|\mathbf{x})$ は、距離関数 d を用いて式(2)で求める。

$$p_\phi(y = k|\mathbf{x}) = \frac{\exp(-d(f_\phi(\mathbf{x}), \mathbf{c}_k))}{\sum_{k'} \exp(-d(f_\phi(\mathbf{x}), \mathbf{c}_{k'}))} \quad (2)$$

また、損失関数は $J(\phi) = -\log p_\phi(y = k|\mathbf{x})$ である。この時、式(2)がSoftmax関数であり合計1であるという性質から、クラス k のプロトタイプに近づけつつ、他のクラスのプロトタイプを遠ざける学習をする。Prototypical NetworkにおけるZero-shot Learningでは、学習していない未知のクラスに対しても、いくつかのサンプルから式(1)で求めたプロトタイプの座標を与えれば、追加学習無しである程度の精度で分類が可能になる。

Learning Common Features between Two Classes in Middle Layer of Prototypical Network

[†]Ryota Sato, Satoru Fujita, Graduate School of Computer and Information Sciences, Hosei University.

3. 2クラス間共通特徴の学習

学習していない未知のクラスを分類するときは、学習したクラスの知識を利用して分類が行われる。ニューラルネットワークが内部で学習した特徴が、未知のクラスに対応しやすい一般化された特徴であるほど、未知のクラスに対する精度は上がるはずである。本研究では、学習した複数のクラスに共通する部分的な特徴は、未知のクラスにも対応できる可能性が高いと考えた。そこで、提案手法では、出力層の1つ手前に中間層を追加し、2クラス間に共通する部分特徴を学習する。中間層のノードは事前学習クラス数 n から2つのクラスを選ぶ組み合わせ $\binom{n}{2}$ とする。損失は、正解ラベルを含むノードを1とするマルチホットベクトルとのクロスエントロピーとする。これによって、中間層のノード一つ一つはクラス i とクラス j が入力された時に1に近づくように学習されるため、ノードがクラス i とクラス j に共通する特徴を学習できる。この中間層の次が多次元座標の出力層になるため、出力する座標は中間層が学習した部分特徴の組み合わせで算出される。

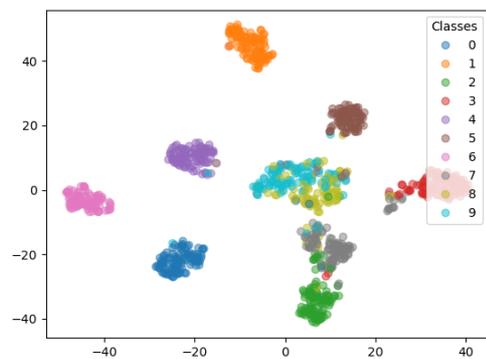


図1. 既存手法の埋め込み空間の分布

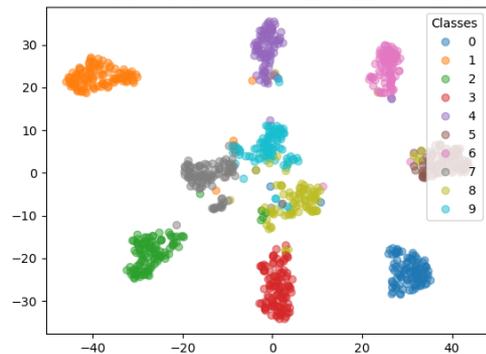


図2. 提案手法の埋め込み空間の分布

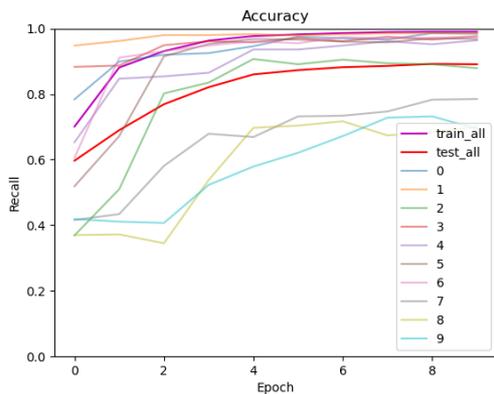


図3. 既存手法の再現率

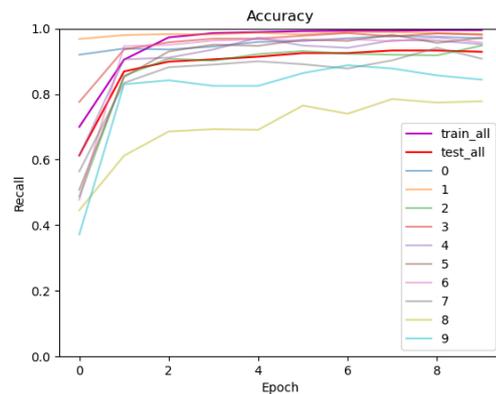


図4. 提案手法の再現率

4. 実験

4.1. 実験方法

分類問題のデータセットでクラス数を減らした状態で訓練を行い、推論時には減らしていたクラスを元に戻した全クラスで精度評価を行う。この実験では、手書き数字画像のデータセットであるMNISTを用いる。訓練時には、7, 8, 9の画像を取り除いた0から6だけの訓練データで学習を行い、テスト時には全てのテストデータを用いて評価する。

ニューラルネットワークのアーキテクチャはResNet [2]をベースとし、出力層では10次元の埋め込み空間上の座標を出力する。出力層の1つ手前の中間層は、訓練データに含まれる7クラスから2クラス選ぶ組み合わせの21ノードとする。

出力の可視化にはt-SNEを用いて、10次元空間の点群データを2次元平面にプロットする。t-SNEは高次元空間上で距離が近いデータを低次元上でも近くに保つことを目的とする次元削減手法である。

分類精度の実験では、テスト時に7, 8, 9を含む全クラスのサンプルからプロトタイプを計算した上で分類を行う。各クラスの精度は再現率によって評価し、全体の精度は全クラスの再現率のマクロ平均で評価する。

4.2. 実験結果

t-SNEで次元削減した既存手法の埋め込み空間への出力の分布を図1、提案手法の出力の分布を図2に示す。未学習クラスの分布について、既存手法では8と9の分布が混ざり合っており、提案手法の方が既存手法よりも分布が別れている。

既存手法の再現率を図3、提案手法の再現率を図4に示す。既存手法では、未学習の7, 8, 9の再現率が0.7程度であるのに対して、提案手法では、7, 8, 9の再現率が0.8程度となっている。

5. 考察

図2に示す提案手法の出力の分布をみると、1, 4, 7の直線で構成される文字が左上に集まっている。一方で、図1に示す既存手法の出力の分布では、分布が隣あっている文字を見ても人間の感覚的にあまり似ているように感じない。これは偶然ではなく、

未学習クラスを変えて複数回実験を行っても、提案手法では1, 4, 7が近くに寄る傾向があった。この結果から、提案手法は2つのクラスに共通する部分的な特徴に注目しているおかげで、似ているパーツで構成される文字が近くに分布するように学習できていると推測する。

6. まとめと今後の課題

本稿ではZero-shot LearningにおいてPrototypical Networkを元に、中間層で2クラス間に共通する部分特徴を学習させる手法を提案した。この手法によって、未学習のクラスに対しても、埋め込み空間上のクラスターをよりはっきりと分割することができた。

今後の課題として、提案手法では、2クラスの組み合わせを学習する層から出力層に繋げているため、出力層では2クラスの組み合わせだけから埋め込み空間上の座標を計算している可能性がある。2クラスの組み合わせを学習した手前の層の特徴量を埋め込み空間への出力にも強く反映させるため、2クラスの組み合わせを学習する層にスキップ構造を追加する案を検討している。

また、MNISTでは各クラスがどんな部分特徴を持っているのかの判断を主観に頼ることになってしまう。Zero-shot Learningの先行研究で用いられているデータセットでは、画像がどんな特徴を持っているのかという属性情報が付加されているものがある。今後の研究では、それらのデータセットを用いて、未知のクラスが既知の部分特徴をうまく活用できているかを詳細に分析する。

参考文献

- [1] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Advances in neural information processing systems*, vol. 30, 2017.
- [2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.