

学習時のみに使用可能な情報を浸透させるニューラルネットワーク

柳元 美玖^{1,a)} 長尾 智晴^{1,b)}

概要: 本稿では「浸透学習」と称する新しいニューラルネットワークの学習法を提案する。浸透学習は、ニューラルネットワークの運用・試験時に使用可能なデータ（モダリティなど）が限定されるシナリオにおいて、訓練時に必要情報を抽出し、「浸透させる」ことによって高い性能を維持するものである。浸透学習は従来の転移学習とは異なり、データ量ではなくデータ次元数の減少への対策となる手法であり、マルチモーダル認識、時系列予測問題、セグメンテーションなど考える応用範囲は広い。本実験では性能評価のためにノイズ加工された画像データセットと5層の単純なニューラルネットワークが用意され、浸透学習した場合とそうでない場合についての精度比較を行った。結果は提案手法の精度が大幅に高いことを示した。

キーワード: ニューラルネットワーク, パターン認識, マルチモーダル認識, 転移学習, 画像認識

Neural Networks Percolating Information Available Only in Training

MIKU YANAGIMOTO^{1,a)} TOMOHARU NAGAO^{1,b)}

Abstract: In this paper, we propose a novel learning method of neural networks called “percolative learning”. The percolative learning is the way to maintain high performance by extracting necessary information when training for “percolation” in a scenario where a subset of data (e.g. modalities) is usable when testing the networks. Unlike existing transfer learnings, the percolative learning is a measure to reduce the number of data dimensions rather than the amount of data. Possible application ranges such as multimodal recognition, time series prediction problem, and segmentation are wide. In this experiment, a noise processed image dataset and a simple networks were prepared for performance evaluation, and the accuracy comparison was carried out in the case of percolative learning and the case without it. The results showed that the accuracy of the proposed method is significantly higher.

Keywords: neural networks, pattern recognition, multimodal recognition, transfer learning, image recognition

1. はじめに

深層学習の隆盛 [1] をきっかけとして、ニューラルネットワークは近年の機械学習において最も注目を集める分野の1つとなった。ただし、深層学習は限られたタスクにおいて人間にも勝る優れた性能を示した [2] が、他のタスク（特に実問題）では依然として認識能力は人間と比べて大幅に劣る場合が多い。それらの原因として最も頻りに指摘されるのがラベル付きデータ量の問題である。一般的に、教師あり深

層学習において高い性能を引き出すためには各モダリティにおける訓練用ラベル付きデータ量を大量に用意する必要があり、その量は実問題において現実的でないことがしばしばある。対して、人間の学習は深層学習ほど大量のデータを要さない。理由は、人間に備わった五感などの感覚器官から得られるマルチモーダルデータ間の関係性を抽出することで、無駄のない情報処理を行うことができているからであると考えられる。このようなマルチモーダル情報処理の考え方は、上記のデータ量の問題を解決するために必要な要素の1つであると言える。

実問題のマルチモーダル認識において、システム運用時にも訓練時と同様のデータ（モダリティ）が得られるとは

¹ 横浜国立大学
Yokohama National University
^{a)} yanagimoto-miku-hb@ynu.jp
^{b)} nagao@ynu.ac.jp

限らない。研究機関が労力をかけて訓練セットとして全てのモダリティのデータを条件を揃えて収集し、システムへの入力とすることはできるが、それによって完成したシステムの運用時にも同様の条件でデータが入力される必要がある。このような必要条件は、システムの実用性を大きく下げたため、実用化は難しい。マルチモーダル情報処理によってデータ量や情報量の問題を解決するためには、異なるモダリティ間が持つ情報を積極的に補いあう技術が求められる。

異なるドメインにおいて得られた知識を当該ドメインにも利用できるようにするという、転移学習の考え方 [11] がこの問題に役に立つ。既存の転移学習研究の多くは、各モダリティに各タスクを割り当て、データ量が十分な一方から得られた知識をデータ量の足りない他方に転移するものである。先のマルチモーダル認識問題に転移学習の考え方を適用する場合、全モダリティに基づき引き出された知識を運用条件を想定して利用可能にすることでであると考える。

本稿では以上に示した訓練時と試験時に使用できるデータが異なる場合に適用可能なニューラルネットの学習方法として、著者らが“浸透学習”と名づけるアルゴリズムを新たに提案する（特許出願済：特願 2017-153613）。

2. 浸透学習

ニューラルネットへの入力として「主データ」と「補助データ」が設定され、訓練時には両データを、試験時には主データのみを使用できる課題を想定する。浸透学習とは、このような課題条件下においても、試験時にデータごと失われる有用な情報を利用できるように、訓練時にネットワークにまるでその情報を“浸透”させるように学習させる手法である。

2.1 構造

浸透学習するニューラルネット（浸透ニューラルネット）の構造的な特色は、補助データがネットワークに与える影響力“非浸透率”をニューラルネットコンポーネントである“スイッチャー”によって外部から制御できる点にある。図 1 に浸透ニューラルネットの構造の概略を示す。スイッチャーは、補助データ x_{aux} と非浸透率 $\alpha_{np} (0 \leq \alpha_{np} \leq 1)$ を入力とし、 $\alpha_{np} x_{aux}$ を出力する関数である。これにより、補助データがニューラルネットに与える影響力を制御できる。例えば、 $\alpha_{np} = 1$ とする場合、補助データの影響力はそのままに連結部分を通じてネットワーク全体に伝わる。非浸透率が小さいほどネットワーク全体に伝搬される補助データ情報の影響率は下がる。 $\alpha_{np} = 0$ とする場合に補助データとしてゼロ行列を入力することと等しくなるため影響力は無くなる。影響力を制限された補助データ $\alpha_{np} x_{aux}$ は、主データ x_{main} と連結され、浸透サブネットに入力される。浸透サブネットの出力を浸透特徴と呼び、これを統

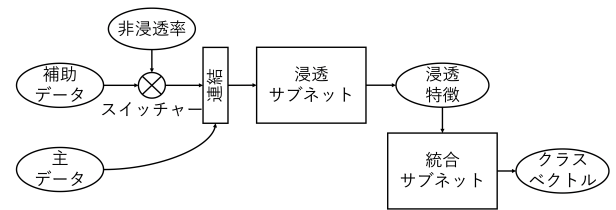


図 1 浸透ニューラルネットの構造。

Fig. 1 Architecture of the percolative neural network.

合サブネットに入力する。統合サブネットの出力はクラスベクトル、すなわち入力データに対して各クラスが正解である確率を要素とするベクトルである。

2.2 学習アルゴリズム

浸透ニューラルネットの学習アルゴリズムを表 1 に示す。アルゴリズムは順番に「事前学習」「浸透学習」「微調整」の 3 フェーズから成り立ち、それらの動作内容を図 2(a)–(c) に示す。事前学習・浸透学習フェーズは必ず実行するが、微調整フェーズはデータとモデルの性質（例えば過学習のしやすさなど）を考慮して実行の有無を決定する。

2.2.1 事前学習

事前学習は提案アルゴリズムにおける最初の学習フェーズである（表 1(a), 図 2(a)）。本研究における事前学習は、 $\alpha_{np} = 1$ に固定した上で、訓練データ（主データと補助データ）を入力し、正解クラスを出力するようにネットワーク全体における重みを最適化する過程を指す。本フェーズの目的は、主データと補助データを等しく使用した場合に最終的に得られる浸透特徴を、続いて行う学習フェーズの教師信号として訓練データごとに記録することにある。一般に、ニューラルネットにおいて、高い（出力層側に近い）層であるほど、与えられたラベル・クラスに関する抽象度の高い特徴を学習していることが知られている [7]。浸透特徴は教師信号として与えられたクラスへの分類に関して一定の抽象度を持った特徴であると言える。浸透学習が有効であるデータは、正しい分類に必要な情報を補助データ及び主データが共通部分としてある程度もっていることを前提としている。この情報の共通部分が浸透特徴にあたり、抽出される浸透特徴の抽象度は浸透サブネットの深さによって調節可能と考えられる。よって補助データを用いて作られた浸透特徴は主データのみによってもある程度復元可能であると考えられることができる。

2.2.2 浸透学習

浸透学習は、記録された浸透特徴を主データのみによって再現させることを目的としたフェーズである（表 1(b), 図 2(b)）。本フェーズでは、非浸透率をエポック（学習回数）ごとに少しずつ減少させつつ、記録された浸透特徴を教師データとした教師あり学習を浸透サブネットのみに対して適用する。非浸透率がほぼゼロになり、かつ記録され

表 1 浸透ニューラルネットの学習アルゴリズム。

Table 1 Learning algorithm of percolative neural networks.

```

/*(a) 事前学習*/
 $\alpha_{np} \leftarrow 1$ 
while(統合サブネットの出力誤差が大):
    全訓練データ・クラスベクトルに対する最適化
    浸透特徴の記録

/*(b) 浸透学習*/
while(浸透サブネットの出力誤差が大 or  $\alpha_{np} \neq 0$ ):
    if  $\alpha_{np} \neq 0$ 
         $\alpha_{np} \leftarrow \alpha'_{np} (0 < \alpha'_{np} < \alpha_{np})$ 
        全訓練データ・浸透特徴に対する最適化

/*(c) 微調整*/
if 微調整する:
    while(統合サブネットの出力誤差が大きい):
        全訓練データ・クラスベクトルに対する最適化
    
```

た浸透特徴と浸透サブネットの最終層の活性ベクトルの誤差が十分に小さくなったと判断された場合に、本フェーズを終了する。このとき、記録された浸透特徴は主データのみによって再現可能になったと考えられる。

2.2.3 微調整

微調整は、浸透学習後に得られた浸透サブネットと統合サブネットの重みを、クラスベクトルに対して再度教師あり学習によって最適化するフェーズである（表 1(c), 図 2(c)）。事前学習との動作の違いは、本フェーズでは非浸透率を 0 に固定している点にある。浸透特徴に対する誤差を完全に 0 にできないことが理由で、実際には浸透学習フェーズ後のクラスベクトルの誤差は浸透学習フェーズ直前よりも上昇し、訓練精度は低下する（節 4 参照）。本フェーズによってクラスベクトルに対する誤差を低下させ、訓練精度を上昇させることができる。ただし、訓練データを与えられたクラスに十分にフィットさせることが、試験精度の上昇につながるとは必ずしも言えない。そのため、データとニューラルネット構造を考慮して、クラスラベルに対して過学習しにくいと判断した場合に本フェーズを実行する。

2.3 関連研究

転移学習は、或るドメイン・タスクのラベル付きデータセットから得られた知識を、それに関連するドメイン・タスクに活用する手法である [11]。本提案手法は、異なるドメインにおいて同じラベルが割り当てられたデータセットが 2 つ以上あり、運用・試験時に取得できるものとできなくなるものがそれぞれ 1 つ以上ある場合に、運用できなくなるデータセット群の知識を訓練中に抽出して運用中にも活用する。上記のように、本研究はこれまでにない新しい転移学習と考えることもできる。他の転移学習との類似性につ

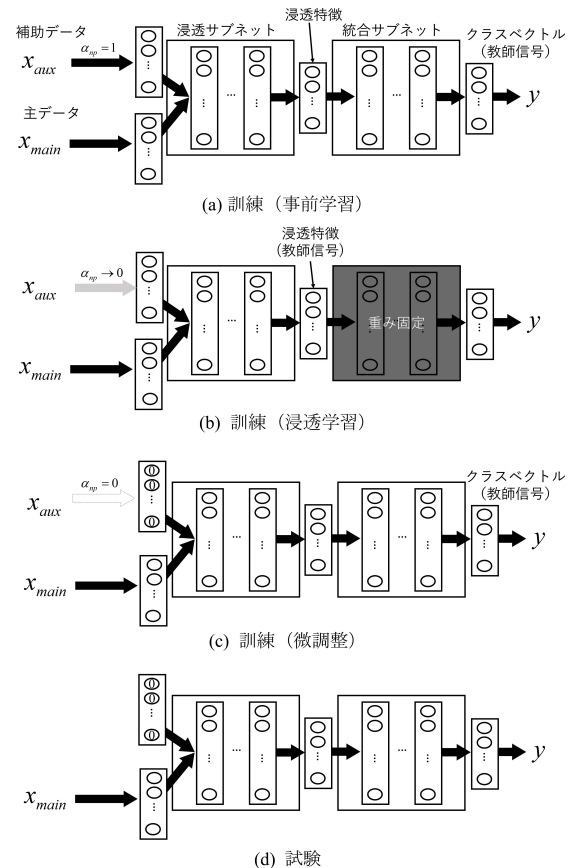


図 2 浸透ニューラルネットの訓練・試験時の動作。

Fig. 2 Functions of percolative neural networks when training and test.

いて、例えば本提案手法が事前学習のフェーズをもつ点が挙げられる。事前学習 (pretraining) には、従来の教師なしと教師あり学習の 2 種類がある。教師なし事前学習として知られるのは、制約付きボルツマンマシンを用いたネットワーク [3][4][5]、あるいはオートエンコーダ [6] などの半教師あり学習における 1 フェーズである。教師あり事前学習としては、転移学習における 1 フェーズが有名である。転移学習には様々なアプローチが存在するが [8][9][10]、中でも画像認識における学習済みモデルを異なるドメインの画像セットに適用する手法は実用的レベルでしばしば用いられる [8]。半教師あり学習を含めた従来の転移学習における事前学習は、ターゲットタスクにおけるデータ量の不足を補うためにある。対して、本提案手法の事前学習は、訓練時と試験時のデータの次元数の違いを補うためにあるため、従来手法とは目的が異なる。

3. 実験設定

3.1 データ

3.1.1 ドメインの選択

実問題における浸透学習の目的は、マルチモーダル認識課題・時系列予測問題などに適用されることである。しかし、本研究では手法の有効性の確認として、上記のような

実用的課題ではなく加工した画像データセットを用いた。その理由は、実問題中のデータセットにおける補助データが主データに対して情報を浸透可能な特性を有しているか否かが明らかでないことにある。浸透可能性とは、主データのみによって浸透特徴がある程度再現できることを指す。浸透可能なデータの必要条件は、補助データ・主データがそれぞれ単独で、与えられたクラスに対して正しく分類されるだけの情報を本来的にある程度有していることである。例えば、マルチモーダル感情認識におけるデータ形式の1つとして脳波を用いる場合、脳波に現れる感情特徴は必ずしも明らかでなく、視覚的に認識可能なものでもない[12]。また、株価変動などの時系列予測問題は、様々な要因が複雑に関係するため、これを正確に解くことは人間にとっても機械にとっても難しい。以上のような人間による認識の困難性をもつデータセットは、与えられたクラスに分類されるための情報を本来的にどの程度有しているかが自明でない。本研究では提案手法の性能評価として人間・機械にとって分類の易しいデータセットを用いることで上記の問題を回避した。主データが持つ情報量を加工によって制限することで、認識課題の難易度を変化させた。

3.1.2 データセット

MNIST データセット [13] を加工したものを実験に用いた。画像データは人間にとって視覚的に特徴を認識可能で、本データセットは数ある画像データセットの中でも情報の有無の程度がわかりやすいことが本データセットを使用する理由である。1 データあたり $28 \times 28 (= 764)$ ピクセルのサイズの手書き数字の白黒画像から成り立ち、訓練セット・試験セット数はそれぞれ 60,000, 10,000 である。

本実験では、訓練における補助データには原データを与えるが、主データとして与えるのは各データ中の $764r_s$ ピクセルの座標がシャッフルされたデータである (図 3(a))。 $r_s (0 \leq r_s \leq 1)$ はシャッフル率を表す。シャッフルされるピクセルの選択は全データに対してランダムであるため、シャッフル率による情報量の制限が可能である (図 4)。

なお、訓練セットは主データと補助データから成り立つが、試験セットは主データのみから成り立つ。図 2(c)(d) が示すように、全ての訓練フェーズを終了した後の試験時には、補助データにいかなる実数値ベクトルを入力しても、 $\alpha_{np} = 0$ によって影響を及ぼさない。図 3(b) の補助データの欄に示すように、全ての補助データの全ピクセルがゼロによって埋められた信号が入力されることと影響は変わらない。

3.1.3 比較手法とデータの与え方

実験において、提案手法に対する比較手法を設定する。比較手法は、提案手法とほぼ同じ構造をもつネットワークを、入力データとクラスベクトルに対して最適化させる従来の学習のみを行う。試験条件は、主データのみから認識を行うことであり、その点は提案・比較手法ともに揃え

	(a) 訓練時	(b) 試験時
主データ		
補助データ		

図 3 訓練・試験時に与えられる主・補助データ例。

Fig. 3 Instances of main and auxiliary data given when training and test respectively.

シャッフル率 r_s	0.25	0.5	0.75	1.0
画像例				

図 4 シャッフル率と画像例。

Fig. 4 Shuffle rate and images.

る。ただし、提案手法と同様の機能 (すなわち訓練時と試験時で使用可能なデータに違いがある場合に性能維持ができる機能) を果たす既存手法がほぼ無いため、提案手法と比較手法の訓練時のデータの与え方は多少異なる。訓練時にネットワークに与える入力データ条件は以下のようになる。

- 提案手法: 主データ, 補助データ, 非浸透率を与える。ここでは非浸透率の与え方を $\alpha_{np} = (1 - decay)^{epoch_p}$ のように設定した。 $decay$ は減衰率であり、予備実験に基づき 0.05 に設定した。 $epoch_p$ は、浸透学習を開始した時点をも 0 回目として数える学習回数である。
- 比較手法 1: 主データのみを与える。通常の学習の結果として得られるネットワークは、学習時に与えられていたデータの条件と同じものを試験時にも与えられなければ有効に機能することはできないことが、本条件を設定した理由である。
- 比較手法 2: 主データと補助データを与える。上記が真だとしても、提案手法の条件と比較して平等でないとも言えることが、本条件を設定した理由である。

3.2 ネットワーク

実験における、提案手法によるネットワークを $y = \text{PercNet}(x_{main}, x_{aux}, \alpha_{np})$ 、比較手法 1, 2 として設定されたネットワークをそれぞれ $y = \text{Net}_1(x_{main})$, $y = \text{Net}_2(x_{main}, x_{aux})$ とする。実験に用いたネットワークは、全結合層 (linear), バッチノーマライゼーション (bn)[14], レクティブファイヤ (relu)[15], ソフトマックス関数 (softmax), スイッチャー (switcher), 連結 (concat) をコンポーネントとして成り立つ。

3.2.1 提案手法のネットワーク

表 2(a) は 5 層の単純なニューラルネットである $\text{PercNet}(x_{main}, x_{aux}, \alpha_{np})$ の構造を表している。スイッチャーは $\text{switcher}(\alpha, x) = \alpha x$ の式で表される。主データ x_{main} と制御された補助データ $\alpha_{np}x_{aux}$ は、 $\text{concat}([x_1], [x_2]) = [x_1, x_2]$ によって連結される。1-4 層は

表 2 実験した浸透ニューラルネットの構造。

Table 2 The architecture of percolative neural network used in the experiments.

(a) PercNet($x_{main}, x_{aux}, \alpha_{np}$)				
層 no.	入力 : サイズ	関数	学習パラメータ : サイズ	出力 : サイズ
	$\alpha_{np}:(1)$			
	$x_{aux}:(764)$	switcher(α_{np}, x_{aux})	無し	$\alpha_{np}x_{aux}:(764)$
	$\alpha_{np}x_{aux}:(764)$			
	$x_{main}:(764)$	concat($\alpha_{np}x_{aux}, x_{main}$)	無し	$x:(1528)$
1	$x:(1528)$	linearunit ($x, 100$)	W, b, γ, β	$x:(100)$
2	$x:(100)$	linearunit ($x, 100$)	W, b, γ, β	$x_{pf}:(100)$
3	$x_{pf}:(100)$	linearunit ($x, 100$)	W, b, γ, β	$x:(100)$
4	$x:(100)$	linearunit ($x, 100$)	W, b, γ, β	$x:(100)$
5	$x:(100)$	linear (x)	$W:(100,10)$	
			$b:(10)$	$x:(10)$
	$x:(10)$	softmax (x)	無し	$y:(10)$

(b) linearunit(x, N)				
層 no.	入力 : サイズ	関数	学習パラメータ : サイズ	出力 : サイズ
1	$x:(D)$	linear(x)	$W:(D, N)$	
			$b:(N)$	$x:(N)$
	$x:(N)$	bn(x)	$\gamma:(1)$	
			$\beta:(1)$	$x:(N)$
	$x:(N)$	relu(x)	無し	$x:(N)$

表 3 実験した比較用ニューラルネットの構造。

Table 3 The architecture of neural networks for comparison used in the experiments.

(a) Net ₁ (x_{main})				
層 no.	入力 : サイズ	関数	学習パラメータ : サイズ	出力 : サイズ
1	$x:(764)$	linearunit ($x, 100$)	W, b, γ, β	$x:(100)$
2	$x:(100)$	linearunit ($x, 100$)	W, b, γ, β	$x_{pf}:(100)$
3	$x_{pf}:(100)$	linearunit ($x, 100$)	W, b, γ, β	$x:(100)$
4	$x:(100)$	linearunit ($x, 100$)	W, b, γ, β	$x:(100)$
5	$x:(100)$	linear (x)	$W:(100,10)$	
			$b:(10)$	$x:(10)$
	$x:(10)$	softmax (x)	無し	$y:(10)$

(b) Net ₂ (x_{main}, x_{aux})				
層 no.	入力 : サイズ	関数	学習パラメータ : サイズ	出力 : サイズ
	$x_{aux}:(764)$			
	$x_{main}:(764)$	concat(x_{aux}, x_{main})	無し	$x:(1528)$
1	$x:(1528)$	linearunit ($x, 100$)	W, b, γ, β	$x:(100)$
2	$x:(100)$	linearunit ($x, 100$)	W, b, γ, β	$x_{pf}:(100)$
3	$x_{pf}:(100)$	linearunit ($x, 100$)	W, b, γ, β	$x:(100)$
4	$x:(100)$	linearunit ($x, 100$)	W, b, γ, β	$x:(100)$
5	$x:(100)$	linear (x)	$W:(100,10)$	
			$b:(10)$	$x:(10)$
	$x:(10)$	softmax (x)	無し	$y:(10)$

表 2(b) で定義された linearunit(x, N) を 4 つ積層したものである。1, 2 層が浸透サブネットにあたり、2 層の出力である x_{pf} は浸透特徴を表す。3-5 層が統合サブネットにあたり、 x_{pf} を入力とし、クラスベクトル y を出力とする。

3.2.2 比較手法のネットワーク

Net₁, Net₂ は共に PercNet のカウンターパートとして設定される。そのため、表 3 と表 2(a) の比較から分かるように構造的な大差はなく、学習アルゴリズムが異なる。なお、PercNet, Net₂ の構造は、Net₁ の構造のカウンターパートとなるように設定された。Net₁ の構造は、必要以上に深く複雑でなく、かつ訓練セット上の学習曲線から安定的に学習が行われていると判断されるものを選んだ。

4. 実験結果と考察

表 4 に各手法において得られた試験・訓練精度を示す。図 5 はその際に得られた学習曲線であり、紫（薄い色）と

表 4 各手法における試験・訓練精度。

Table 4 Accuracies on test and training sets of each method.

(a) $r_s = 0.5$		
手法	試験精度 [%]	訓練精度 [%]
比較 1	83.90	99.97
比較 2	16.97	99.97
提案	86.75	92.43
提案 (+ 微調整)	86.59	99.94

(b) $r_s = 0.75$		
手法	試験精度 [%]	訓練精度 [%]
比較 1	53.17	99.67
比較 2	12.11	99.98
提案	64.04	73.04
提案 (+ 微調整)	58.36	99.49

緑（濃い色）の曲線がそれぞれ訓練精度と試験精度である。提案手法は、微調整をした場合としない場合について記載している。

図 5 の提案手法の学習曲線に着目すると、比較手法のそれとは形状が大きく異なっているが、理由は学習アルゴリズムの切り替えにある。0-1,000 エポックにおいて事前学習が、1,000-11,000 エポックにおいて浸透学習が、11,000-13,000 エポックにおいて微調整が行われている。事前学習は比較手法 2 と等しく、主データと補助データが平等にネットワークに入力され、最適化が行われる。よって補助データに大幅に依存した最適化が行われることが原因で、高い訓練精度と非常に低い試験精度が得られる。浸透学習の開始によって、訓練精度と試験精度はそれぞれ急激に低下または上昇する。訓練精度の低下は、入力データの変化によって学習パラメータセットが得るべき解が全く異なるものになるためである。試験精度の上昇は、訓練時の入力データ条件が試験時のそれに近づいたためである。浸透学習フェーズの終了時に訓練精度が 100% より大幅に下回った理由は、記録された浸透特徴に対する誤差を十分に小さくすることができなかったためである。原因は 2 つ考えられ、1 つは情報量を減少された主データの浸透可能性の限界、もう 1 つは浸透特徴の良質さ、すなわち浸透サブネットの適切性である。後者の問題について、本実験で使用されたニューラルネットの主となるコンポーネントは全結合 (linear) 層であり、ネットワーク自体が過学習しやすい傾向にあったと考えられる。過学習が起きている判断については、図 5 の比較手法の試験・訓練曲線に着目して、訓練精度が最大となるより先のエポックで試験精度が最大になっている点から推測される。

表 4 において提案手法と比較手法の試験精度を比較すると、いずれの場合においても提案手法が一貫して高く、シャッフル率が高いほどその差は大きい傾向もわかる。表 4 における比較手法 2 の試験精度は、他のいずれの手法と

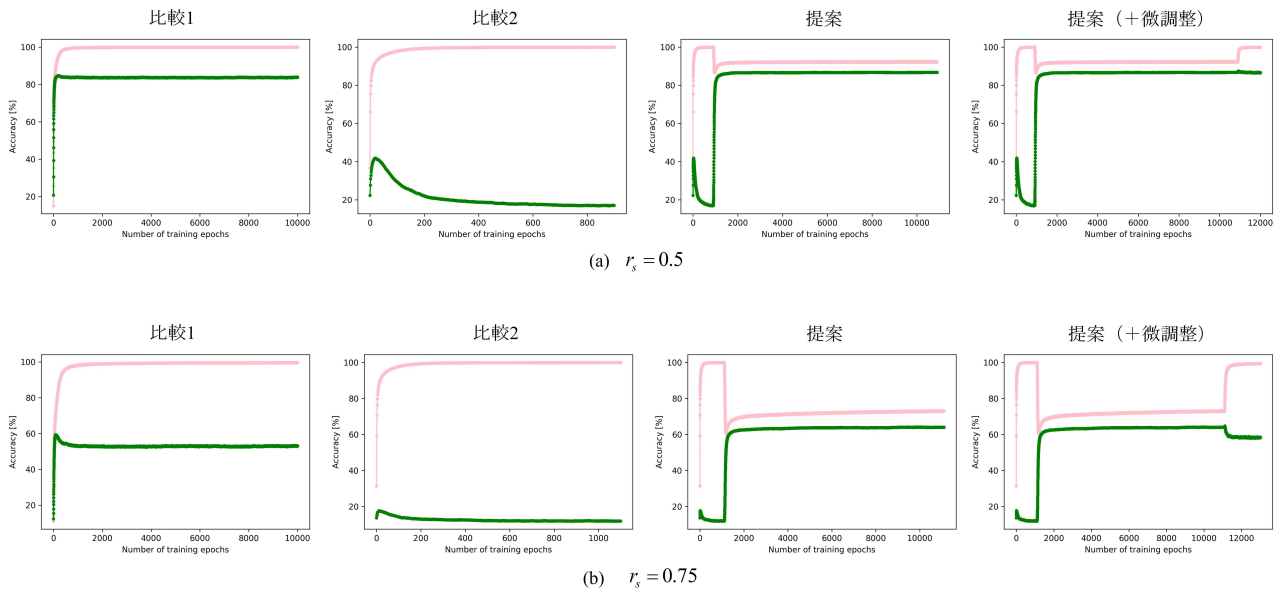


図 5 各手法における学習曲線。
Fig. 5 Learning curves of each method.

比較しても大幅に精度が低く、これは訓練時に比較手法 2 の訓練精度が最も高い理由は、学習を終了するまで補助データを制約なく入力され続けることにある。提案手法の微調整をしない場合とした場合について、本実験で設定された条件下では微調整をしない方が試験精度が高かった。ただ、微調整をする方が良いかどうかはモデルとデータに依存するため、しない方が良いとは一概に言えない。

5. 結論

本稿では、試験時にデータごと失われる有用な情報を訓練時に浸透させるように学習させる手法である浸透学習を提案した。実験に用意したのは、加工を施した画像データセットと、全結合層を主とする単純なニューラルネットであり、浸透学習をさせる場合とさせない場合について精度比較を行った。結果は浸透学習の精度が比較手法のそれを大幅に上回ることを示した。

参考文献

- [1] Krizhevsky, A., Sutskever, I., and Hinton, G.E. :ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems* 25, pp. 1097–1105, (2012).
- [2] He, K., Zhang, X., Ren, S., et al. :Deep residual learning for image recognition, *arXiv preprint arXiv:1512.03385*, (2015).
- [3] Hinton, G.E. and Salakhutdinov, R.R. :Reducing the dimensionality of data with neural networks, *Science*, vol. 313 no. 5786, pp. 504–507, (2006).
- [4] Hinton, G. E., Osindero, S. and Teh, Y.-W. :A fast learning algorithm for deep belief nets. *Neural Comp.* vol. 18, no. 7, pp. 1527–1554, (2006).
- [5] Salakhutdinov, R., and Hinton, G.E. :Deep Boltzmann

- machines, *AISTATS*, pp. 448–455, (2009).
- [6] Hinton, G.E., and Salakhutdinov, R.R. :Reducing the dimensionality of data with neural networks, *Science* vol. 313, no. 5786, pp. 504–507, (2006).
- [7] LeCun, Y., Bengio, Y., and Hinton, G.E. :Deep learning, *Nature*, no. 521, pp. 436–444, (2015).
- [8] Jia, Y., Shelhamer, E., et al. :Caffe: Convolutional architecture for fast feature embedding, *arXiv preprint, arXiv:1408.5093*, (2014).
- [9] Makhzani, A., Shlens, J., Jaitly, J., et al. :Adversarial auto-encoders, *International Conference on Learning Representations*, (2016).
- [10] Ganin, Y., Lempitsky, V. : Unsupervised domain adaptation by backpropagation, *International Conference on Machine Learning (ICML)*, (2015).
- [11] Pan, S.J., and Yang, Q. :A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.*, vol. 22 no. 10, pp. 1345–135 (2010).
- [12] Yanagimoto, M., Sugimoto, C., and Nagao, T. :Frequency filter networks for EEG-based recognition, *International Conference on Systems, Man, and Cybernetics (SMC)*, (2017).
- [13] LeCun, Y., and Cortes, C. The MNIST database of handwritten digits.
- [14] Ioffe, S., and Szegedy, C. :Batch normalization: Accelerating deep network training by reducing internal covariate shift, *International Conference on Machine Learning (ICML)*, pp. 448–456, (2015).
- [15] Nair, V., and Hinton, G.E. :Rectified linear units improve restricted Boltzmann machines, *International Conference on Machine Learning (ICML)*, pp. 807–814, (2010).