2V-01

人間からニューラルネットワークへの知識転移に関する基礎研究

鈴木　一誠†　　ピトヨ・ハルトノ‡
中京大学　工学研究科†　　中京大学　工学部‡

1.　はじめに

現在、Deep Neural Network(DNN)の実世界問題への応用が盛んに行われる。DNNの学習には構造化されたデータが必要であるため、データ化の難しい人間の常識、主観や経験などをニューラルネットワーク(NN)に反映させることは難しい。そのため、NNから他のNNに知識の転移に関する研究[1]は多くあるが、人からNNに知識を転移する手段はまだ少ない。

そこで、本研究では2次元の位相的な中間層を持つNN、Restricted Radial Basis Function Network (rRBF)[2]を用いて、人間の常識や経験をNNに転移させる手法を提案する。提案手法では、人間が常識、経験に基づいてNNの2次元の中間層を手動で組織化することで、NNを初期化することができる。人間の初期化によって、その人間の特性が後のNNの学習に反映されることを期待する。本研究では、人間の特性転移によって初期化されたNNにエントロピーによる解析を行い、その結果を報告する。

2.　人間の特性転移が可能なNN

本研究で用いるrRBFは、2次元の中間層を持つ階層型NNであり、教師付き学習を実行することができる。rRBFの中間層の組織化はSelf-Organizing Maps (SOM)[3]とは異なり、入力のラベル(context)を反映するためこの中間層をcontext-relevant self-organizing map(CRSOM)という。CRSOMは2次元であり、マップとして可視化することができるため、rRBFを次元圧縮アルゴリズムや可視化による入出力関係を直観的に理解できるNNとして用いる[4,5]。

本研究では、2次元のCRSOMの位相的な性質を用いて人間からNNへ特性転移を可能とする手法の提案をする。初めに、人間が経験、好み、知識を基に手動で中間層の組織化をする。具体的には、人間は類似する入力同士を2次元の中間層上で近くに配置し、類似しない入力同士を中間層上で遠くに配置する。人間による組織化が行われた後に、rRBFの教師付き学習を実行する。

組織化の仕方が個人の常識、経験や主観によって異なるため、rRBFにそれを組織化する人間の特性を転移でき、学習後のNNにもその特性が反映されると考える。本研究で用いるrRBFの概要を図1に示す。ここでは、中間層と出力層の間に畳み込み層を置き、組織化の概要を示す。

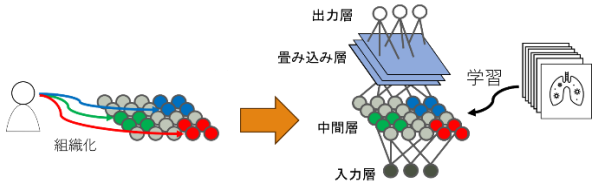


図1　rRBFの概要

Research on the knowledge transfer from humans to neural network
†Issei Suzuki
Graduate School of Engineering Chukyo University
‡Pitoyo Hartono
School of Engineering Chukyo University

2.1 rRBFの学習

人間による組織化後、rRBFは以下のように学習する。

rRBFに対し、$X(t) \in R^d$を時間tにおいての入力とし、それに対するBest Matching Unit (BMU)、winを式(1)に示す。

$$win = argmin\left\|X(t) - W_{ij}\right\|^2 \tag{1}$$

式(1)の$W_{ij} \in R^d$は中間層上の(i,j)に位置するニューロンに対応する参照ベクトルである。BMUの決定後、(i,j)に位置する中間ニューロンの出力O_{ij}^hを式(2)に示す。

$$O_{ij}^h(t) = e^{-I_{ij}^h(t)}\sigma(win,(i,j),t) \tag{2}$$
$$I_{ij}^h(t) = \left\|X(t) - W_{ij}\right\|^2$$

次に近傍関数$\sigma(win,(i,j,t)$を式(3)に示す。

$$\sigma(win,(i,j),t) = e^{-\frac{dist(win,(i,j))}{s(t)}} \tag{3}$$
$$s(t) = s_0\left(\frac{S_{end}}{s_0}\right)^{\frac{t}{t_{end}}}$$

式(3)の$dist(win,(i,j))$,は、中間層上でのBMUと(i,j)に位置するニューロンのデカルト距離を示す。s_0, s_{end}は、それぞれ学習の始まりと終わりの近傍半径を示す定数である。また、t_{end}は学習回数である。このような近傍関数を用いることで、学習回数により近傍半径は変化する。

次の畳み込み層では、フィルタ$F(n,c,r,s)$、バイアスbとした場合、(i,j)に位置するニューロンの出力$O_{ij}^{conv}(t)$を以下に示す。

$$O_{ij}^{conv}(t) = \sum_{c=1}^{C}\sum_{r=0}^{R-1}\sum_{s=0}^{S-1} O_{(i+s)(j+r)}^h(t)\cdot F(n,c,r,s) + b \tag{4}$$

式(4)のcは入力のチャンネル数、nは出力のチャンネル数、rはフィルタの高さ、sはフィルタの幅を表す。k番目の出力ニューロンO_kを式(5)で計算する。

$$O_k(t) = f\left(\sum_j v_{(ij)k}(t) O_{ij}^{conv}(t) - \theta_k(t)\right) \tag{5}$$

$$f(x) = \frac{1}{1+e^{-x}} \tag{6}$$

式(5)の$v_{(ij)k}$は(i,j)に位置する畳み込み層のニューロンとk番目の出力ニューロンとの間の重み、θ_kはk番目のニューロンのバイアスである。$f(x)$は式(6)で示すSigmoid関数である。

ここで、Loss関数を以下のように定義し、T_kは教師信号のk目の要素を示す。

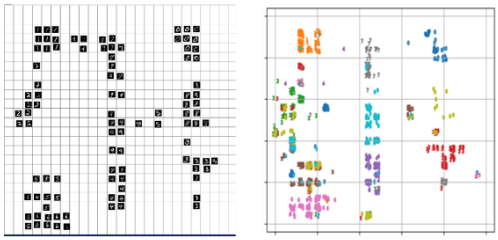
$$Loss = \frac{1}{K}\sum_{k=1}^{K}\left(T_K(t) - O_k(t)\right)^2 \tag{7}$$

学習では、重み、参照ベクトル、フィルタパラメータの更新を行う。実装ではPythonの自動微分機能を用いる。

3. 実験

実験データは、「0」～「9」の画像データ、MNIST[6]を用いる。MNIST は、手書によって作成されたため、個人の特性が反映するからである

実験では、「0」～「9」のをそれぞれ 10 枚、合計 100 枚を人間が個人の主観で中間層に配置することで組織化を行う。人間が行った初期化の例を図 2(a)に示す。ここでは、ラベルに関わらず、類似する入力同士を 2 次元の中間層上で近くに配置し、類似しない入力同士を中間層上で遠くに配置する。図 2(b)は学習後の CRSOM を示す。また、色は教師ラベルの違いを表す。



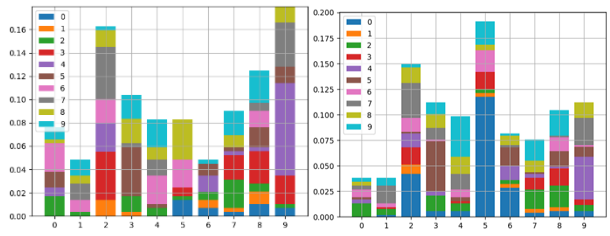
(a) (b)

図 2 人の初期化による rRBF 学習

図 2(a)と図 2(b) より学習後の CRSOM が人間による初期化を反映することがわかる。これは、人間の知識や主観が rRBF に反映することを意味する。

また、各ラベルに対するランダム初期化のエラーバーを図 3(a)、人の初期化のエラーバーを図 3 (b)に示す。このグラフの横軸は手文字画像の ground truth であり、縦軸は他の文字への誤認識率を示す。ここでは、i 番目の rRBF が学習後に、文字jの誤認識率を$p_i(j)$、文字jを文字kとして誤認識率を$pf_i(j,k)$とする。$N_i(j,k)$ を文字j を文字k と誤認識した数とする。

$$p_i(j)=\frac{\sum_{k\neq j=0}^{9}N_i(j,k)}{\sum_{j=0}^{9}\sum_{k\neq j=0}^{9}N_i(j,k)},\qquad pf_i(j,k)=\frac{N_i(j,k)}{\sum_{k\neq j=0}^{9}N_i(j,k)}\ (8)$$



(a) (b)

図 3 初期化に対する rRBF の誤認識率

人間の知識転移の実験では、5 人の被験者がそれぞれ異なる rRBF を初期化した後に、それらの rRBF の教師付学習を実装した。比較として、ランダムに初期化した 5 つの rRBF の学習も行った。ランダムに初期化した rRBF を random1~5、5 人の被験者を person1~5 とし、それらの rRBF の差異を評価する。h番目の rRBF とi番目の rRBF の差異の評価方法は、エラー全体の分布に注目する $coarse(h,i)$ (式 9)とエラー詳細な内容に注目する $detail(h,i)$ (式 10)である。

$$coarse(h,i)=\frac{1}{2}\sum_{j=0}^{9}\{-pf_i(j)\log pf_h(j)-pf_h(j)\log pf_i(j)\}\ (9)$$

$$detail(h,i)=\frac{1}{20}\sum_{j=0}^{9}\sum_{k\neq j-0}^{9}\{-p_i(j,k)\log p_h(j,k)-p_h(j,k)\log p_i(j,k)\}\ (10)$$

式(9)、(10) に基づいて、異なる rRBF の差異を計算した。random の平均と person の平均を表 1 に示す。

表 1 rRBF の特性の差異

	random の平均	person の平均
coarse	4.236	3.143
detail	4.250	3.484
| coarse-detail |	0.014	0.341

coarse では、person の平均値が random の平均値よりも小さい。これは、数字の類似性に関する人の常識には共通点が多く、常識転移を行わない複数のニューラルネットワークの方に大きな多様性が生じる。一方、(coarse-detail)の値において person が大きい。これは、人間の常識によって初期化された各ニューラルネットワークが異なるエラーの発生の仕方をしたことを意味する。つまり、人間が初期化した rRBF がその人間の特性を反映すると考えられる。

4. まとめ

本研究では人間の知識を NN に事前知識として転移できる手法を提案した。ここでは、手書き文字を用いて基礎実験を行った。実験の結果からは人間が NN に転移した常識が、その後の NN の特性に影響を与えることが分かった。このことは、データからだけでなく、人間から学ぶ NN を学習させることができ、人間と AI の新しい関係を確立できると考える。

今後は、データ化することが困難な問題にこの学習方法を適応することを検討する。例えば、職人的な技術や経験や主観的を必要とするスキルの NN の学習を試みる。

謝辞

本研究はローム株式会社との共同研究による成果である。

参考文献

[1] 神嶌敏弘, “転移学習” 人工知能学会誌, Vol.25, No.4 (2010).
[2] P. Hartono, et al, “Learning-Regulated Context Relevant Topographical Map,” in IEEE Transactions on Neural Networks and Learning Systems, Vol.26, No.10 (2015).
[3] T. Kohonen, “Self-organized formation of topologically correct feature maps.” *Biological cybernetics* 43.1 (1982).
[4] P. Hartono, “Classification and dimensional reduction using restricted radial basis function networks”, Neural Computing & Applications, Vol.30, No.3 (2018).
[5] P. Hartono, “Mixing autoencoder with classifier: conceptual data visualization”, *IEEE Access*, Vol.8 (2020).
[6] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-based learning applied to document recognition,” in Proceedings of the IEEE, Vol.86, No.11 (1998).