F-002

ニューラルネットワークへの人間の主観転移に関する研究
Research on human subjective transfer to neural networks

鈴木 一誠† ピトヨ・ハルトノ‡
Issei Suzuki Pitoyo Hartono

1. はじめに

現在、Deep Neural Network(DNN) の様々な実世界問題への応用が盛んに行われる。しかし、DNN の学習には、大量の訓練データと膨大な学習時間を必要とし、必ずしも実用的でない場合もある。訓練データの多さと学習時間の問題を解決するための 1 つの方法としてある問題領域で学習済みのニューラルネットワーク(NN)の一部を類似した問題領域で学習を行う他の NN への転移の研究［1,2］が行われる。NN から他の NN に知識の転移を行う研究は多くあるが、人から NN に知識 を転移する手段はまだない。人から NN への一般常識や基礎知識を転移することができれば、NN の学習の困難さを緩和できるだけでなく、人間の意思決定の特性を反映する NN を学習することが可能となり、更に新しい人間と AI の関係を確立できると考える。

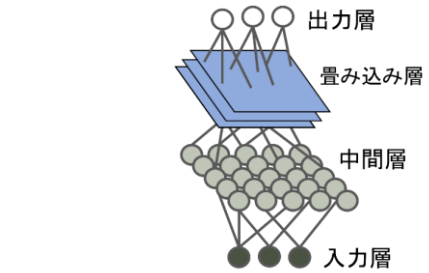
そこで、本研究では人間の知識、経験や主観を NN へ転移する手法を提案する。本研究は過去に提案した位相的な 2 次元の中間層をもつ階層型 NN、Restricted Radial Basis Function Network (rRBF)[3]を用いる。rRBF の中間層には多次元入力の位相構造が組織化されるため、人間はその構造を可視化することで直観的に理解することができる。また、中間層が 2 次元であるため、人間が経験や知識に基づいてその中間層を手動で組織化できる。この組織化を行うことで、人間の知識や経験が NN の事前知識として転移され、その後の学習に反映される。本研究では手書き文字認識をタスクとして、人間の知識の転移に関して実験を行い、NN を学習する。異なる人間が組織した学習後の NN を解析し、個人差が NN の差に反映されるかに関して解析を行い、その結果を本論文で報告する。

2. 人間の知識転移可能な NN

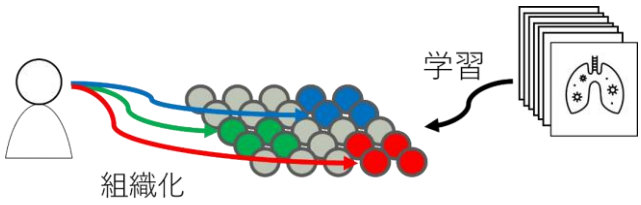
本研究で用いる rRBF を図 1(a) に示す。rRBF は、２次元の中間層を持つ階層型 NN であり、教師付き学習を実行することができる。rRBF の中間層の組織化は Self-Organizing Maps (SOM)[4]とは異なり、入力のラベル(context)を反映するためこの中間層を context-relevant self-organizing map(CRSOM)という。CRSOM は 2 次元であり、マップとして可視化することができるため、rRBF を次元圧縮アルゴリズムや可視化による入出力関係を直観的に理解できる NN として用いる[5,6]。

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 への知識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。つまり、ここでは rRBF を人間の知識や主観的な考えによって組織化する。

組織化の仕方が個人の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 の構造を図 1 (a)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、組織化の概要を図 1(b)に示す。



(a)



(b)

図 1　rRBF の構造と中間層の組織化

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d次元の時間t においての入力ベクトルとし、それに対する Best Matching Unit (BMU)、win を式(1)で定義する。

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

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)で計算する。

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

次に近傍関数$\sigma(win,ij,t)$を式(3)と定義する。

† 中京大学大学院　工学研究科　電気電子工学専攻
Department of Electrical and Electronic Engineering, Graduate School of Engineering Chukyo University

‡ 中京大学　工学部
School of Engineering Chukyo University

$$\sigma(win, ij, t) = e^{-\frac{dist(win,ij)}{s(t)}} \quad (3)$$

$$s(t) = s_0 \left(\frac{s_{end}}{s_0}\right)^{\frac{t}{t_{end}}}$$

式(3)の$dist(win, ij)$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。このように、近傍関数を定義することで、学習回数により近傍半径は変化する。

本研究では、中間層と出力層の間に畳み込み層を追加する。ここでは、フィルタ$F(n, c, r, s)$、バイアスb とした場合、畳み込み層で(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。

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

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。

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

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

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。

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

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。また、実装では Python の自動微分機能を用いる。

3. 初期実験

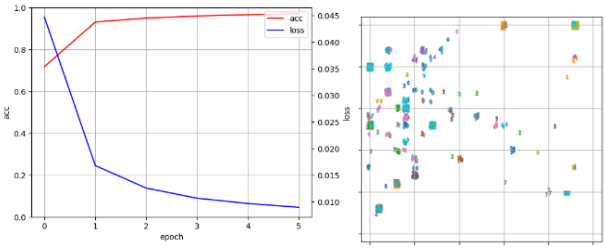
初期実験では、28×28 ピクセルの手書き文字「0」～「9」の画像データ、MNIST[7]を用いる。MNIST を用いる理由は、手書によって作成されたデータであるため、個人の特性が反映するからである。また、見る人によっても字の類似性と相違性に関する感覚が異なる。そのため、NN への人間の特性の転移の実験に適している。

初めに、転移学習を行わず、中間層をランダムに初期化し、rRBF 学習を行った結果を図 2 に示す。図 2(a)は学習過程を示す。

図 2(b)は学習後の CRSOM で、全入力に対する BMU を示す。ただし、異なる入力が同一の BMU を持つ場合、小さな乱数を用いてその位置をずらす。また、色とラベルの対応は表 1 に示す。

表 1 ：ラベルと色の対応

ラベル	色
0	青
1	橙
2	緑
3	赤
4	紫
5	茶色
6	ピンク
7	グレー
8	黄色
9	水色



(a) (b)

図 2　ランダム初期化の rRBF 学習

図 2(a)からこの rRBF はうまく学習できることがわかったが、テストデータに対する正解率は 8.92%と汎化能力が低い。原因は、図 2(b)より同一の BMU が異なるクラスを持つ複数の入力に対して反応することにある。

この初期実験から、中間層での組織化と汎化能力の間に強い関係があることがわかる。さらに、ランダムに初期化された rRBF の学習は困難であることもわかった。

4. 人間からの転移学習実験

実験では、判別が難しい文字を含めた「0」～「9」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。組織化に使用する一部の画像を図 3 に示す。

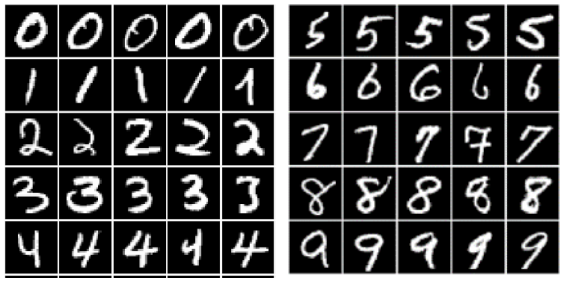
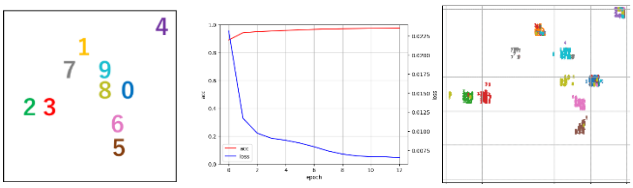


図 3　組織化に使用する一部の画像

人間が行った組織化の例を図 4 (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。そうすることで、中間層上での組織化が配置した個人の主観を表す。人間の組織化に使用しない 59900 枚を

訓練データとして rRBF の学習に用い、1 万枚のテストデータを用いて rRBF の汎化能力の評価を行う。図 4(a)で組織化した rRBF の学習過程を図 4(b)に示し、図 4(c)は学習後の CRSOM を示す。ただし、異なる入力が同一の BMU を持つ場合、小さな乱数を用いてその位置をずらす。



(a) (b) (c)

図 4　人の知識による組織化の rRBF 学習

図 4(a)と図 4(c) より学習後の CRSOM の組織が人間による初期組織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。このとき、テストデータに対する正解率は 96.69%であることから、人間からの知識の転移がその後の学習に寄与することがわかる。

次に、同様の実験を複数の人間に対して行った。ここでは、個人個人で異なる組織化を行い、それぞれを用いて rRBF を学習する。それらの accuracy(正解率)を図 5 に示す。ここでは、ランダムに組織化した rRBF を random、5 人の被験者を person1~5 で表示する。

また、ランダム初期化または人間によって組織化された rRBF の学習後の特性を図 6(a)～(f)に示す。図の上段の左側は学習前の CRSOM を示し、上段の右側は学習後の CRSOM を示す。下段は各ラベルに対するエラーバーを示す。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。また、グラフの色は表 1 に示す関係と同様である。

エラー率を式(8)で計算する。ここでは、i 番目の被験者が組織化した rRBF が学習後に、文字jを文字k として誤認識率を$p_i(j,k)$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。

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

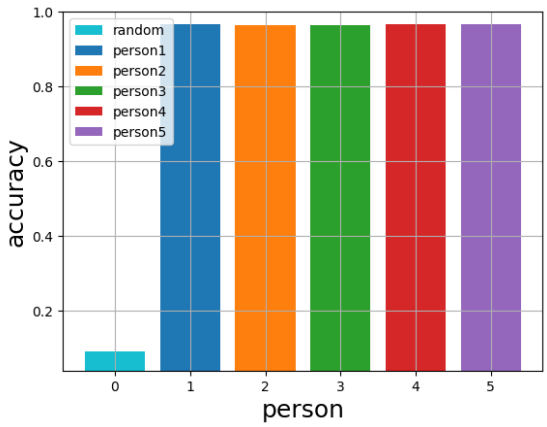
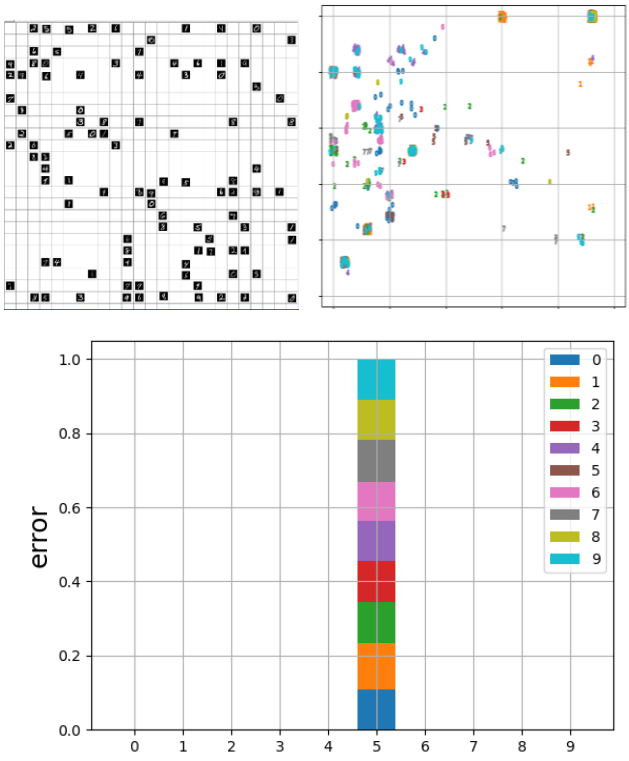
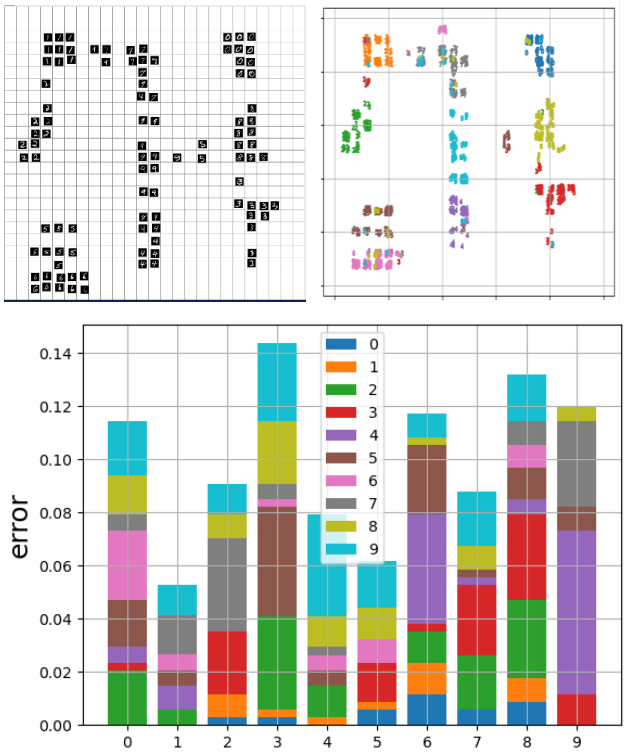


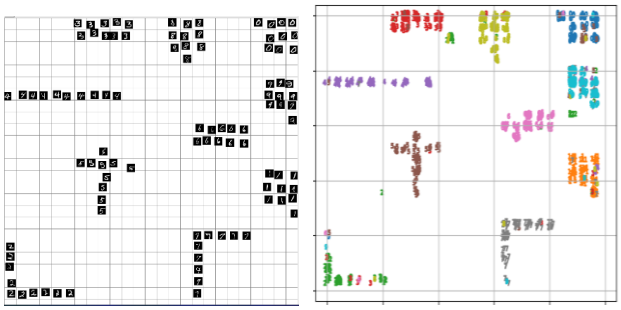
図 5　rRBF の正解率

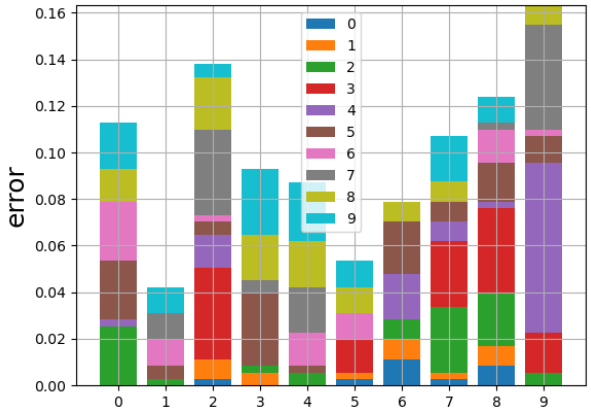


(a)　random

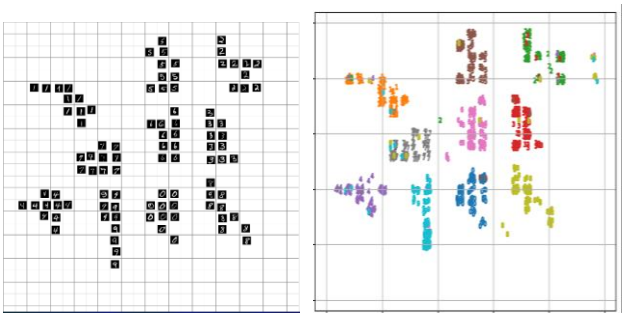


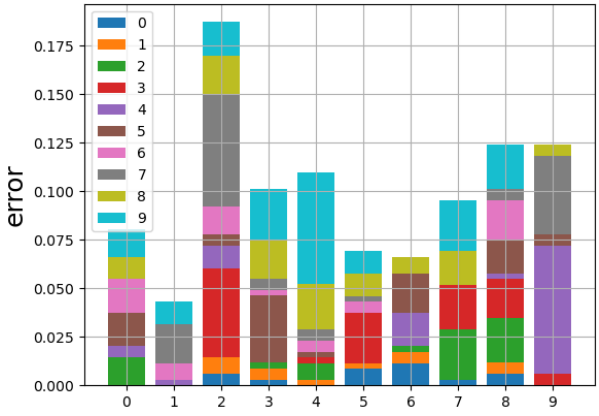
(b) person 1



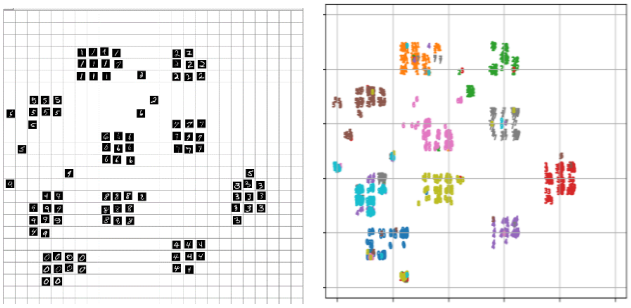


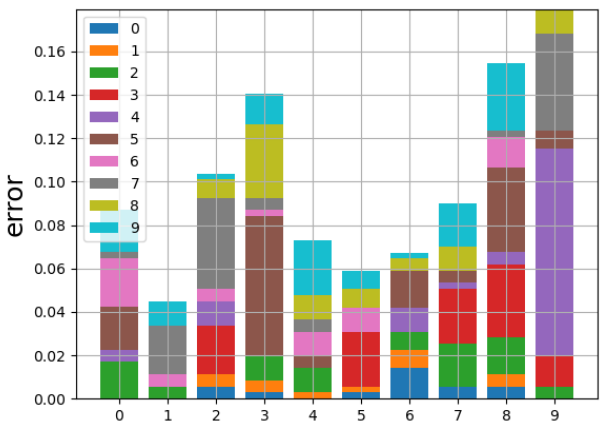
(c) person 2



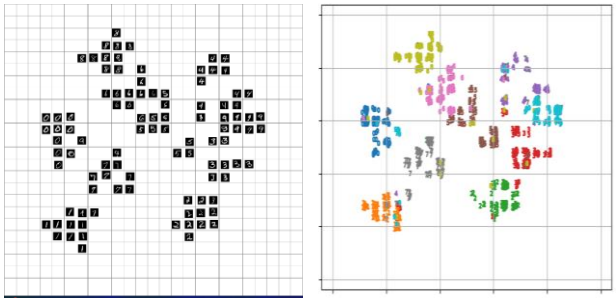


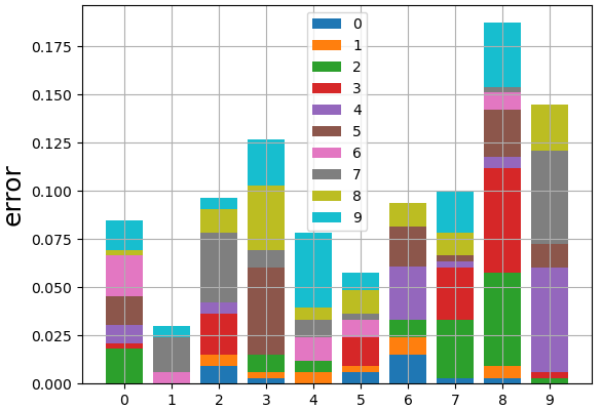
(e) person 4





(d) person 3





(f) person 5

図 6　組織化に対する rRBF の誤認識率

次に、異なる人間によって初期化された rRBF の差異を評価する。h番目の rRBF とi番目の rRBF の差異を評価するために、$dif(h,i)$ を式(9)に定義する。

$$dif(h,i)=\frac{1}{90}\sum_{j=0}^{9}\sum_{k\neq j=0}^{9}(p_h(j,k)-p_i(j,k))^2 \quad (9)$$

その結果を表２に表す。

表 2：rRBF の特性の差異

	person1	person2	person3	person4	person5
random	1.220×10^{-3}	1.248×10^{-3}	1.292×10^{-3}	1.236×10^{-3}	1.262×10^{-3}
person1		4.197×10^{-5}	6.275×10^{-5}	5.309×10^{-5}	4.276×10^{-5}
person2			4.945×10^{-5}	3.886×10^{-5}	4.729×10^{-5}
person3				6.266×10^{-5}	4.197×10^{-5}
person4					5.929×10^{-5}
person5					

図 6 からは異なる人によって組織化された rRBF は異なる特性を持つことがわかる。これは、学習後 CRSOM の可視化から明らかである。また、エラーグラフから各々の rRBF が異なるエラー特性を持つことも明らかである。この差異をさらに表 2 で定量化することで、人間による転移された事前知識がその後の rRBF の学習に影響を与えることを明確にした。

5. まとめ

本研究では人間の知識をニューラルネットワーク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ニューラルネットワークに転移した事前知識がその後のニューラルネットワークの特性に影響を与えることが分かった。この結果から、学習データに含まれない特徴を転移学習によってニューラルネットワークに学習できることが分かった。このことは、データからだけでなく、人間から学ぶニューラルネットワーク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などのニューラルネットワークの学習を試みる。

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] 中山 英樹, “深層畳み込みニューラルネットワークによる画像特徴抽出と 転移学習”,信学技報, Vol.115, No.146 (2015).
[2] 神嶌敏弘, “転移学習” 人工知能学会誌, Vol.25, No.4 (2010).
[3] P. Hartono, et al, “Learning-Regulated Context Relevant Topographical Map,” in IEEE Transactions on Neural Networks and Learning Systems, Vol.26, No.10 (2015).
[4] T. Kohonen, “Self-organized formation of topologically correct feature maps.” *Biological cybernetics* 43.1 (1982).
[5] P. Hartono, “Classification and dimensional reduction using restricted radial basis function networks”, Neural Computing & Applications, Vol.30, No.3 (2018).
[6] P. Hartono, “Mixing autoencoder with classifier: conceptual data visualization”, *IEEE Access*, Vol.8 (2020).
[7] 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).