

画像の深層学習のためのクラスタ傾向分析を用いた 視覚的分析手法

佐藤 信¹

概要: 本稿では、画像の深層学習のためのクラスタ傾向分析を用いた視覚的な分析手法を提案する。提案手法では、学習過程での深層ネットワークの各層の重みおよび出力の変化を把握するために、クラスタ傾向分析をおこなう。特に、クラスタ傾向分析に用いるデータの前処理が分析結果に与える影響について検討をおこなう。学習ラベルを用いず各層の重みおよび出力のみを用いて分析をおこなうことから、教師あり学習だけでなく教師なし学習の学習過程の可視化にも用いることが可能である。提案手法は、学習過程の大局的な分析に適した手法である。

Visual Analysis Method with Cluster Tendency Analysis for Image Deep Learning

MAKOTO SATOH¹

Abstract: This paper presents a visual analysis method with cluster tendency analysis for image deep learning. In the method, the cluster tendency analysis is developed for figuring out the change of the each layer weights and the outputs of deep networks. Effects of preprocessing of analyzing data are discussed. The method can be used for visualizing learning process of unsupervised learning as well as supervised learning, using only the each layer weights and the outputs of deep networks, while not using training labels. The method is suitable for comprehensive analysis of machine learning process.

1. はじめに

本稿では、画像の深層学習のための視覚的な分析手法を提案する。Hinton 等 [6] により、画像の学習において深層学習を用いることにより高い学習性能が得られることが示されて以来、深層学習に関する研究が盛んにおこなわれている。深層学習は、データ集合から特徴を抽出する性能に優れた手法である。しかし一方では、学習モデルが複雑であること、および大量の学習データを必要とすることなどから、モデルの性能を引き出した学習をおこなうことが難しい手法であるともいえる。そのため、深層学習のモデルの理解を助けることを目的として、可視化手法についての研究が盛んにおこなわれている。

提案する可視化手法の特徴を次に示す。

- 深層ネットワークの各層の出力についてクラスタ傾向分析をおこなうことにより、大局的に学習過程を分析することが可能である。
- クラスタ傾向分析の前処理として、分析の対象となるデータを選択することにより、計算量の削減をおこなう。

これ以降の構成について、簡単に説明する。2 節では、関連研究と提案手法との関係について述べる。そして、3 節では、画像の深層学習のためのクラスタ傾向分析を用いた視覚的な分析手法を提案する。実験結果および検討について 4 節で述べる。そして最後に、5 節で本稿のまとめと今後について述べる。

2. 関連研究との比較

2.1 深層学習の可視化

コンピュータおよびインターネットの普及に伴い、大量

¹ 岩手大学
Iwate University, Ueda, Iwate 020-8551, Japan

のデータを取り扱うための手法の必要性が増加しているといえる。それらの手法のなかで、大量のデータがもつ特徴を視覚的に把握するためのデータ可視化手法に関する研究が活発におこなわれている [2], [10].

機械学習の分野においては、学習モデルの入力データ、および出力として得られる分析結果などが可視化の対象となるのはもちろんであるが、それに加えて、学習モデルの学習過程、および学習により得られるパラメータなども可視化の対象となる。そして、機械学習により得られる学習モデルを可視化する目的は、学習モデルを理解、診断、および洗練化することであるといえる [11].

機械学習の重要な分野である深層学習においても、可視化手法に関する研究が活発に行われている。それらの研究の多くは、深層ネットワークが学習により獲得する各層の重みなどの、学習モデルの個々のパラメータを可視化しようとするものであるか、または、畳み込み深層ネットワークの畳み込み層が出力する特徴マップなどの、学習モデルにある入力を与えた場合の各層の出力などを可視化しようとするものである。

一方、少数ではあるが、深層ネットワークの各層の学習状態の関係を可視化しようとする研究もおこなわれている [9]. 提案手法も、深層ネットワークの各層の関係を分析することを目的とする手法である。その特徴は、学習モデルの各層の出力がどのようなクラスタにより構成されるのかを可視化することにより、学習モデルの大局的な特徴を視覚的に捉えて分析することである。それに加えて、ある着目したデータが各層にどのように写像されるのかを可視化することにより、個々のデータが各層にどのような関係で写像されるのかについて視覚的に分析することが可能である。

2.2 クラスタ傾向分析

データ集合のクラスタ傾向分析とは、あるデータ集合がどのようなクラスタにより構成されるのかについて分析をおこなうための手法である。例えば、予めハイパーパラメータとしてデータを構成するクラスタ数の概要を与えなければならない機械学習のアルゴリズムを用いる場合に、クラスタ数の見積もりをするためにクラスタ傾向分析が用いられる。

視覚的にクラスタ傾向を分析するための手法として、VAT(Visual Analysis for Cluster Tendency Assessment) [1] および iVAT(Improved Visual Analysis for Cluster Tendency Assessment) [13] がある。データ集合 S が与えられているとして、始めに、VAT のアルゴリズムの概要を示す。

Step 1 S に含まれる全てのデータの組について、各組のデータの非類似性を計算する。その値を行列の要素の値とする非類似性行列 D を作成する [5]. D の i 行 j

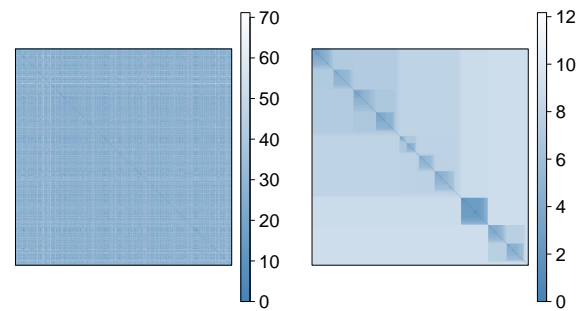


図 1 MNIST の分類結果の iVAT によるクラスタ傾向分析の例
Fig. 1 Examples of cluster tendency analysis of MNIST classification results with iVAT. Left: Visualization of a dissimilarity matrix. Right: The result of iVAT using the dissimilarity matrix in the left panel.

列の要素の値は、 S のデータ i と j との非類似性を表わす。

Step 2 D の行および列を、隣接する行および列に対応するデータの非類似性の差が小さくなるように入れ替える。これにより得られる行列を、行列 R とする。

Step 3 R をグラフィックス表示する。表示では、 R の各要素の値として格納されている非類似性の値を、画像の輝度に対応させる。

iVAT は、VAT を改良した手法である。VAT では、非類似性の計算にデータ空間上のデータ点間のユークリッド距離などを用いる。しかし、そのような直接的な距離を用いただけでは、十分にクラスタ傾向を把握できないような場合が存在する。そこでデータ点間のパスを考慮するように改良したのが、iVAT である。iVAT では、非類似性を計算しようとする 2 つのデータ点を端点として、一方の端点から他のデータ点を経緯して他方の端点に到達するパスを考える。そして、データ点間の距離の計算では、パスを構成するデータ点間の距離が小さい場合にはパスの端点となるデータ点どうしの非類似性の値が小さくなるような距離を用いる。そのような距離を用いて作成した非類似性行列について VAT をおこなうのが iVAT である。

図 1 は、深層ネットワークのある層の出力のクラスタ傾向を、iVAT を用いて可視化した例である。ここで用いた深層学習モデルは、4.1 節の実験で用いた LeNet5 [7] を基本とした学習モデルである。その深層ネットワークを MNIST [8] の訓練データ集合により訓練した学習モデルを用いて、MNIST のテストデータ集合から取り出した画像 1000 枚を分類した結果 (最後の層の 10 次元の出力) について、iVAT を用いてクラスタ傾向を分析した例である。図 1(左) が非類似性行列を可視化したものであり、図 1(右) が iVAT の結果である。このように、クラスタ傾向を、低い輝度の矩形領域として表現した画像が得られる。この例からは、分析をおこなったデータは、約 10 クラスタから構

成されていることが分かる。

提案手法では、深層ネットワークの各層の出力データのクラスタ傾向を視覚的に分析するために、iVATを用いる。

3. 画像の深層学習のためのクラスタ傾向分析を用いた視覚的分析手法

3.1 アルゴリズムの概要

提案アルゴリズムの概要を、アルゴリズム1に示す。アルゴリズムは、次の3段階により構成される。

Step 1 訓練データを用いて学習モデルを訓練する。このとき、予め決めておいたエポックでの学習モデルを保存する。

Step 2 **Step 1**で保存した学習モデルに、分析用データ集合に含まれる全データを入力として与える。そして、分析の対象となるレイヤからの出力を保存する。

Step 3 分析用データ集合から、着目するデータを選択する。そして始めに、**Step 2**で保存した各レイヤの出力について、iVATを用いてクラスタ傾向を可視化した画像を作成する。次に、作成した画像に、着目するデータとして選択したデータの写像点を表示する。ここでの写像点とは、iVATによる再配置後の非類似性行列の行または列のうちで、着目するデータが対応する行または列のことである。

3.2 分析データ選択のための前処理

iVATは、分析の対象となるデータのデータ数の増加に伴い、計算量が急激に増加するアルゴリズムである。そのため、分析の品質を保ちながら、可能な限り少数のデータにより分析をおこなうことが重要である。分析対象となるデータを減少させるための前処理を次に示す。

Step 1 アルゴリズム1の**Step 1**を実行する。

Step 2 アルゴリズム1の**Step 2**および**3**をある1つの学習モデルのレイヤについて実行する。

Step 3 **Step 2**でのiVATで作成した再配置後の非類似性行列から等間隔*i*で行を選択する。そして、選択した行に対応するデータを縮小した分析用データ集合の候補とする。縮小した分析用データ集合の候補を用いて、学習モデルのいくつかのレイヤについてアルゴリズム1の**Step 3**を実行する。

Step 4 分析のための計算量および分析の品質が満足できるものであれば、候補のデータ集合を正式の縮小した分析用データ集合として、アルゴリズム1の**Step 2**および**3**を必要な回数だけ実行する。一方、分析のための計算量および分析の品質が不十分である場合には、**Step 2**に戻り、データ選択のための間隔*i*を変更して、**Step 2**, ..., **4**を繰り返す。

Algorithm 1 Visual analysis method with cluster tendency analysis for image deep learning

Step 1

Prepare training dataset \mathcal{T} .

Prepare analyzing deep learning model.

Set maximum epoch for training models to N .

Set trained model saving epochs in set \mathcal{E} .

$n \leftarrow 0$

while $n < N$ **do**

Train the model for one epoch with the training dataset \mathcal{T} .

if n exists in \mathcal{E} **then**

Save currently trained model parameters into trained model set \mathcal{M} .

$n \leftarrow n + 1$

end if

end while

Step 2

Prepare analyzing dataset \mathcal{V} .

for all the trained models in the trained model set \mathcal{M} **do**

Test the current trained model with the analyzing dataset \mathcal{V} .

Save the each layer outputs of the model, in layer output set \mathcal{L} .

end for

Step 3

Select dataset \mathcal{A} from \mathcal{V} , which will be investigated in terms of the layer mapping of the trained model.

Analyze the cluster tendency of the layer output set \mathcal{L} with iVAT.

Mark the corresponding rows or columns of the iVAT images, that are corresponding to the data in \mathcal{A} .

Plot and save the resultant images.

3.3 教師あり学習または教師なし学習への適用

提案のアルゴリズムを、教師あり学習または教師なし学習に適用する方法について述べる。アルゴリズム1の**Step 1**および**Step 2**については、教師あり学習と教師なし学習とのどちらについても、適用の方法に違いはない。**Step 3**については、教師あり学習では、ラベルと予測値との差を基準として用いることにより、分析において着目するデータを選択することが可能である。しかし、ラベルを用いずに、分析において着目するデータを選択することが可能な場合もありえるので、そのような場合には、教師あり学習と教師なし学習とは、提案手法を適用する場合の差はないといえる。

4. 実装と結果の検討

4.1 実装および実験方法の概要

3節で述べたアルゴリズムを用いて、深層畳み込みネットワークを用いて画像の学習をおこない、そして、学習した深層ネットワークの学習状態を可視化することにより視覚的な分析をおこなった。

表 1 ネットワーク・アーキテクチャの概要
 Table 1 Overview of network architecture

notation	modules	description
C1	convolution	input channel number: 1, output channel number: 6, convolution kernel size: 5
R1	ReLU	
P1	max-pooling	kernel size: 2, step size: 2
C2	convolution	input channel number: 6, output channel number: 16, convolution kernel size: 5
R2	ReLU	
P2	max-pooling	kernel size: 2, step size: 2
F3	fully-connected	input dimension: $16 \times 4 \times 4$, output dimension: 120
R3	ReLU	
F4	fully-connected	input dimension: 120, output dimension: 84
R4	ReLU	
F5	fully-connected	input dimension: 84, output dimension: 10
S5	log softmax	

深層ネットワークのアーキテクチャには、LeNet5 [7] を基本としたアーキテクチャを用いた。実験に用いたアーキテクチャの概要を表 1 に示す。実際の構成では、データ augmentation を目的として入力画像にノイズを付加するためのモジュール、および、モジュール間で受け渡すテンソルデータの形状を変更するためのモジュールが、表 1 の構成に加わる。深層ネットワークの実装には、Torch [3] を用いた。

そして、訓練した深層ネットワークを視覚的に分析するためのアルゴリズムを、R 言語 [12] により実装をおこなった。iVAT を用いたクラスタ傾向分析のためには R 言語のパッケージ seriation [4] などを用いた。

また、学習データおよび分析用データには、手書き数字の画像データセットである MNIST [8] を用いた。

実験の手順の概要を次に示す。

Step 1 MNIST の訓練データ集合および訓練ラベル集合を用いて、深層畳み込みネットワークを訓練した。予め決めておいたエポックにおいて、学習モデルを保存した。

Step 2 分析用データとして MNIST のテストデータ集合を、Step 1 で保存した学習モデルの入力として与えた。そして、学習モデルを構成する各モジュールの出力を保存した。保存したデータについて、iVAT を用いてクラスタ傾向の可視化をおこない、分析画像を保存した。

Step 3 表 1 の S5 層の出力として得られる各クラスの対数尤度を基準として用いることにより、分析用のテストデータから着目するデータを選択した。そして、Step 2 において iVAT によりクラスタ傾向を可視化した画像に、選択したデータの写像点を表示した。

4.2 着目する画像と各層出力のクラスタ傾向との関係

分析用の MNIST 画像を深層ネットワークの入力として、各モジュールの出力を分析した。その結果を、図 2 に示す。図の各列は左から右に、訓練エポック 1,2,10 および 100 における学習モデルを用いた結果である。各行は上から下に、モジュール P1, R3, R4 および S5 の出力データを分析したものである。各セルには、iVAT を用いたクラスタ傾向分析により得られる画像、および、着目する分析用データが iVAT により写像される行および列の位置を示す色付けされた線分が描かれている。線分の色は数字のクラスに対応している。また、着目する分析用データの選択には、分析用データを分類した場合にモジュール S5 の出力として得られる対数尤度を基準として用いた。具体的には、分析用データの各クラスについて、対数尤度の大きいほうから 10 データを用いた。

図 3 は、数字 1 および 2 について、正しく分類された画像および誤って分類された画像が写像される iVAT 分析画像の行および列の位置を、色付けされた線分により描いたものである。図の各列は左から右に、訓練エポックが 1,2,10 および 100 における学習モデルを用いた結果である。各行は上から下に、モジュール P1, R3, R4 および S5 の出力データを分析したものである。各セルには、iVAT を用いたクラスタ傾向分析により得られる画像、および、数字 1 および 2 の分析用データが iVAT により写像された画像の行および列の位置を示す色付けされた線分が描かれている。線分の色は、各数字の分類結果の正誤に対応している。

4.3 前処理としてのデータ選択の効果

選択をおこなったデータ集合についての分析結果を、図 4 に示す。図の各列は左から右に、訓練エポック 1,2,10 および 100 における学習モデルを用いた結果である。各行は上から下に、分析用データ数が 1000, 500, 100 および 50 の

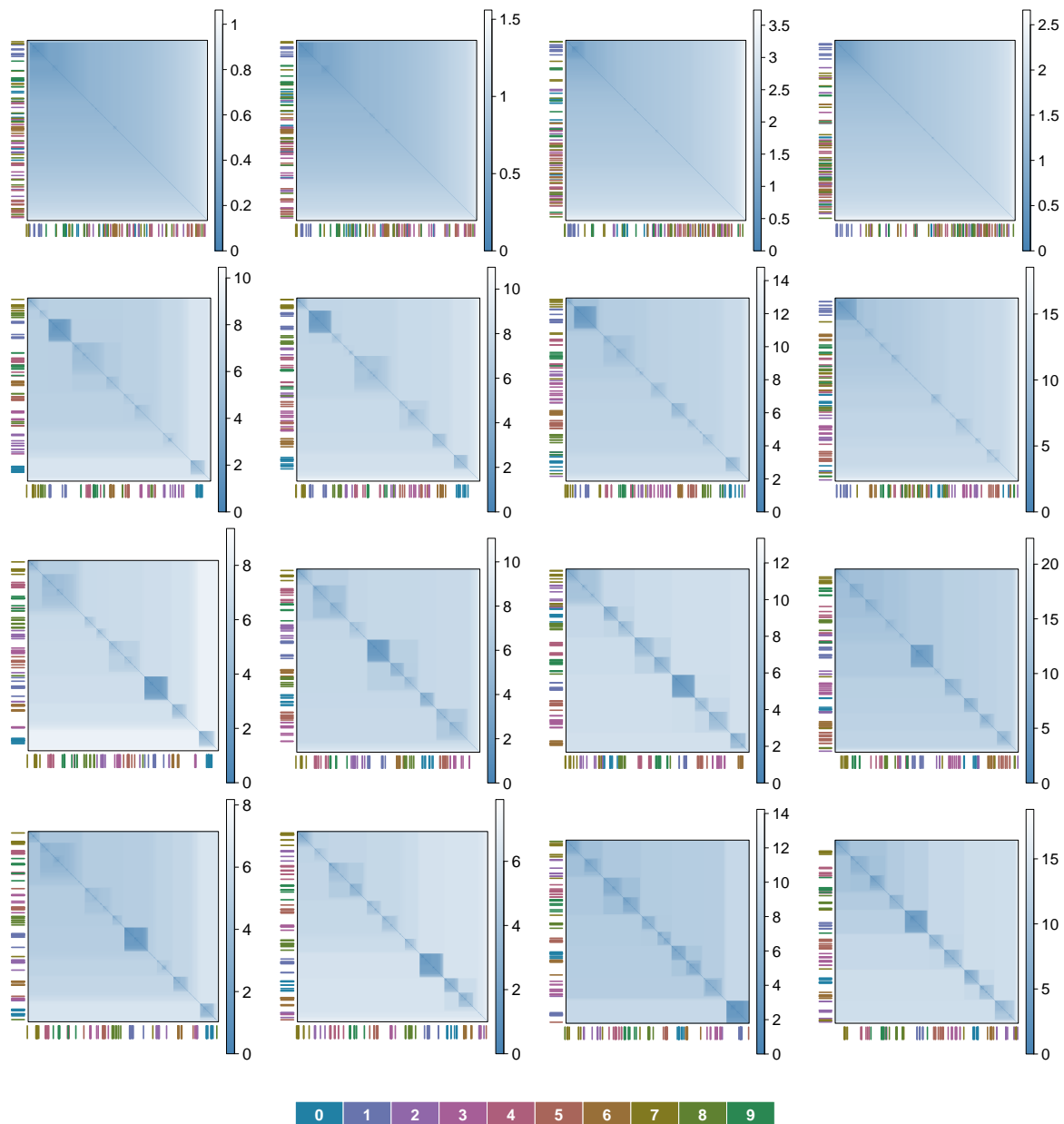


図 2 MNIST 画像データを用いた訓練過程における深層ネットワークのモジュール出力のクラスラ傾向

Fig. 2 Cluster tendency of the module outputs of deep network during the course of training with MNIST images: Left to right: training epoch 1, 2, 10 and 100. Top to bottom: module P1, R3, R4 and S5.

場合の分析結果である。各セルには、iVAT を用いたクラスラ傾向分析により得られる画像、および、着目する分析用データが iVAT により写像される行および列の位置を示す色付けされた線分が描かれている。線分の色は数字のクラスに対応している。

データを選択した方法は、次のとおりである。始めに、学習済みの深層ネットワークの入力としてデータサイズ 1000 の分析用データを与え、その場合のモジュール S5 の出力を iVAT によりクラスラ傾向分析した。iVAT により

得られる再配置行列の行の順序づけにあわせて、分析用データを並べ替えた。そして、並べ替えた分析用データから等間隔でデータを選択した。

4.4 検討

図 2 の最下行は、深層ネットワークの全結合層の最後の層であるモジュール S5 の出力の分析結果である。訓練エポック数の増加に伴い、10 個の低輝度の矩形領域が明確になる様子が分かる。このことから、訓練エポック数の増加

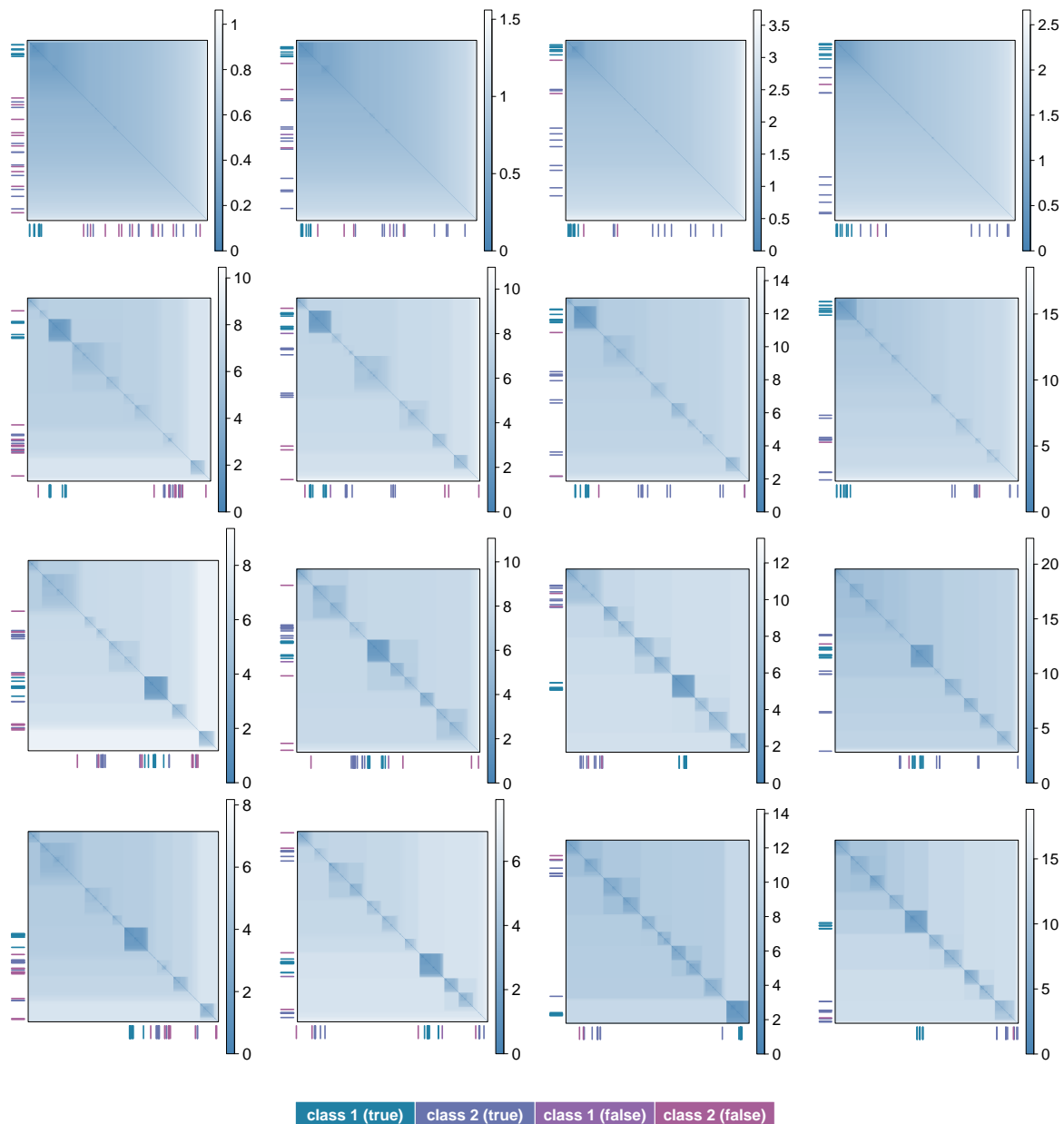


図 3 クラスタ傾向分析での 1 および 2 の数字画像の写像位置

Fig. 3 Mapping positions of number symbol image 1 and 2 with cluster tendency analysis. Left to right: training epoch 1, 2, 10 and 100. Top to bottom: module P1, R3, R4 and S5.

に伴い、分析データが 10 のクラスから構成されるデータであることを学習できていることが分かる。

次に、着目するデータとして各クラスから選択したデータと低輝度の矩形領域との対応について確認する。訓練エポック数の増加に伴い、各矩形領域には同一色の短い線分が対応するようになるのが分かる。短い線分の色は分析データのクラスに対応しているので、訓練エポック数の増加にともない、正しいクラス分けが可能であるように学習モデルが訓練されていることが分かる。

図 2 の第 2 および 3 行は、全結合層の最初の 2 つの層

であるモジュール R3 および R4 の出力である。前述のモジュール S5 についての分析結果とあわせてこれらの行に示す図について分析すると、全結合層の後ろの層になるにしたいが、データのクラス構成を明確に分類するように学習しているといえる。

一方、畳み込み層の複数のチャンネルのうちの 1 つのチャンネルの出力を分析した結果である図 2 の第 1 行では、数字 1 が、一つのクラスを構成するように写像されていることが分かる。しかし、他の数字についてはクラスを構成していないことが分かる。この点については、今後の検討

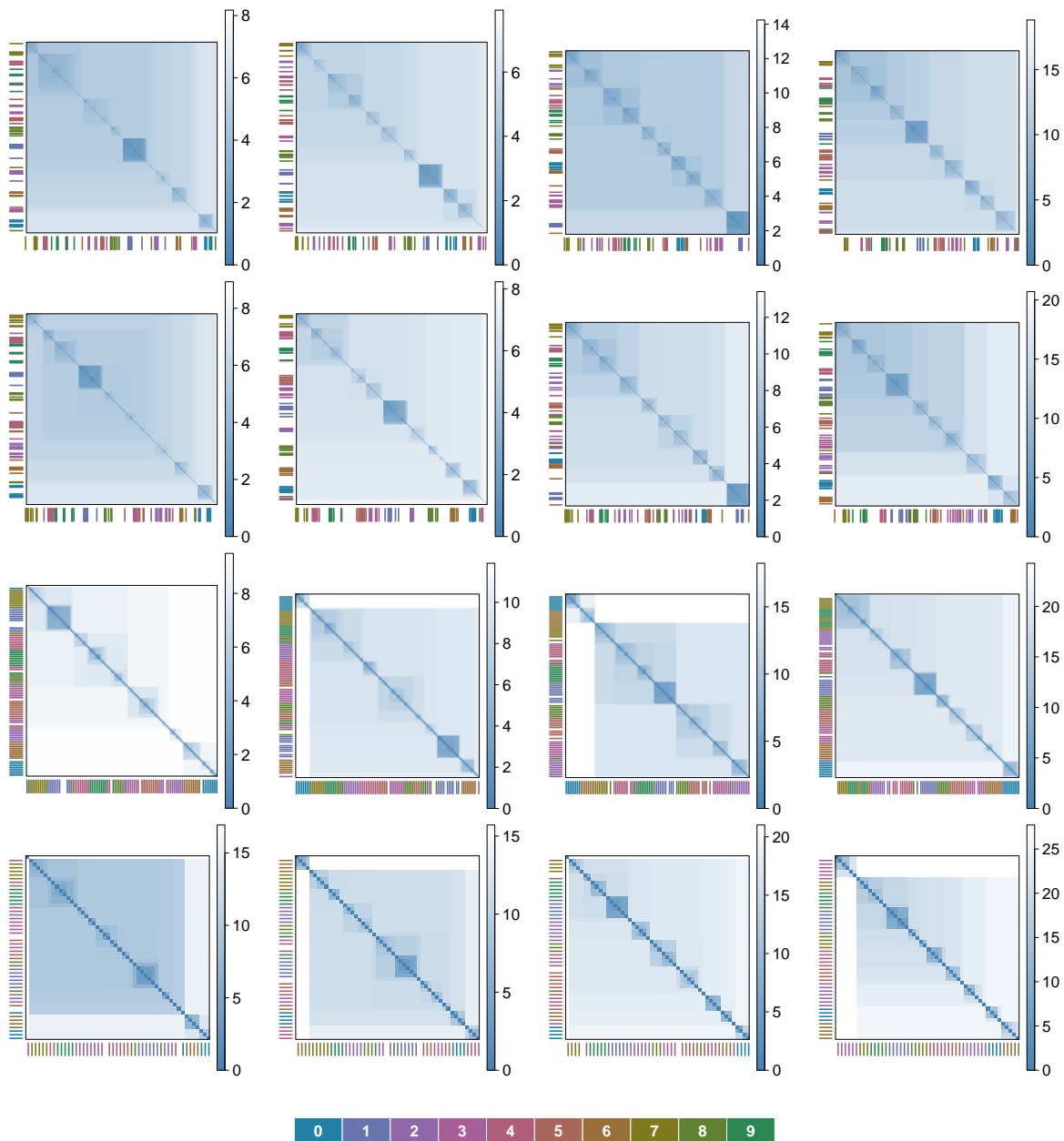


図 4 選択をおこなったデータ集合についてのクラスタ傾向分析

Fig. 4 Cluster tendency analysis on selected datasets. Left to right: training epoch 1, 2, 10 and 100. Top to bottom: selected dataset size 1000, 500, 100 and 50.

課題である。

図 3 からは、誤って分類されたデータとモジュール出力のクラスタとの対応関係(写像)が分かる。図 3 の最下行に着目すると、誤ってあるクラスとして分類されたデータは、そのクラスのクラスタの端のほうに写像されていることが分かる。また、図 3 の各行に着目すると、そのような傾向は、訓練エポックの増加に伴い明確になることが分かる。

図 4 からは、提案手法によりデータ数を 1000 から 500 まで減少させてもほぼ同様の結果が得られることが分かる。しかし、データ数を 100 または 50 まで減少させた場合には、同様の結果を得ることができないことが分かる。

5. おわりに

画像の深層学習のための視覚的な分析手法を提案した。提案した手法の特徴は、クラスタ傾向分析を用いることにより、深層ネットワークの学習過程を大局的に分析することが可能な点である。また、分析の対象となるデータのデータ数を減少させるための手法を示した。

畳み込み深層ネットワークを例として、提案手法の有効性を示した。着目するデータが、各層の出力データ集合のクラスタに対してどのような関係で写像されるのかを可視化することにより、深層学習の学習過程の視覚的な分析を

おこなった。また、提案手法を用いて、分析の対象となるデータ数を減少させても、ある程度、学習過程の視覚的な分析が可能であることを示した。

今後の課題には、既提案手法の他の深層ネットワーク・アーキテクチャへの適用、および、よりわかりやすい可視化手法に関する研究などを挙げるができる。

3-642-13657-3.5 (2010).

参考文献

- [1] Bezdek, J. C. and Hathaway, R. J.: VAT: a tool for visual assessment of (cluster) tendency, *Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on*, Vol. 3, pp. 2225–2230 (online), DOI: 10.1109/IJCNN.2002.1007487 (2002).
- [2] Brodbeck, D., Mazza, R. and Lalanne, D.: *Interactive Visualization - A Survey*, pp. 27–46, Springer Berlin Heidelberg (2009).
- [3] Collobert, R., Kavukcuoglu, K. and Farabet, C.: Torch7: A Matlab-like Environment for Machine Learning, *BigLearn, NIPS Workshop* (2011).
- [4] Hahsler, M., Buchta, C. and Hornik, K.: Infrastructure for seriation. R package version 1.2-1, <https://cran.r-project.org/web/packages/seriation/index.html> (2016).
- [5] Hahsler, M. and Hornik, K.: Dissimilarity Plots: A Visual Exploration Tool for Partitional Clustering, *Journal of Computational and Graphical Statistics*, Vol. 20, No. 2, pp. 335–354 (online), DOI: 10.1198/jcgs.2010.09139 (2011).
- [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., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324 (online), DOI: 10.1109/5.726791 (1998).
- [8] LeCun, Y. and Cortes, C.: MNIST handwritten digit database, (online), available from (<http://yann.lecun.com/exdb/mnist/>) (2010).
- [9] Liu, M., Shi, J., Li, Z., Li, C., Zhu, J. and Liu, S.: Towards Better Analysis of Deep Convolutional Neural Networks, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 23, No. 1, pp. 91–100 (online), DOI: 10.1109/TVCG.2016.2598831 (2017).
- [10] Liu, S., Cui, W., Wu, Y. and Liu, M.: A Survey on Information Visualization: Recent Advances and Challenges, *Vis. Comput.*, Vol. 30, No. 12, pp. 1373–1393 (online), DOI: 10.1007/s00371-013-0892-3 (2014).
- [11] Liu, S., Wang, X., Liu, M. and Zhu, J.: Towards better analysis of machine learning models: A visual analytics perspective, *Visual Informatics*, Vol. 1, No. 1, pp. 48 – 56 (online), DOI: <http://dx.doi.org/10.1016/j.visinf.2017.01.006> (2017).
- [12] R Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2016).
- [13] Wang, L., Nguyen, U. T. V., Bezdek, J. C., Leckie, C. A. and Ramamohanarao, K.: iVAT and aVAT: Enhanced Visual Analysis for Cluster Tendency Assessment, *Proceedings of the 14th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining - Volume Part I, PAKDD'10*, Berlin, Heidelberg, Springer-Verlag, pp. 16–27 (online), DOI: 10.1007/978-