

# ディープラーニングの学習評定への応用に関する研究

A study on application of deep learning to five rank evaluation

白石 正 人

曾根崎 幹 大

野 方 健 治

Masato SHIRAISHI

Mikihiko SONEZAKI

Kenji NOGATA

技術教育研究ユニット

糸島市立前原中学校

技術教育研究ユニット

(令和5年9月4日受付, 令和5年12月22日受理)

## 抄 録

ペーパーテストの結果だけでなく多様な視点から学習を評定することが求められている。また、単元毎の学習評価については、観点別評価を行うものの最終的な評定に関しては学校の裁量に委ねられており、統一した規準等は存在しない。一方、AI（人工知能）のディープラーニングは、中間層を複数有し、画像認識等に大きな成果を挙げており、この技術を用いて多様な成績データから評定（5段階評定あるいは、秀、優、良、可、不可の判定）を推定することが可能かどうか検証した。成績評価については、学校現場でExcelを用いることが多く、このマクロ機能を提供するVBAを用いてディープラーニングの評定推定プログラムを試作した。学部2科目の成績データを用いてこの評定推定プログラムに適用したところ、3層のニューラルネットワークでは76.7%、4層のディープラーニングでは83.7%の最大正答率が得られた。パラメータ等の最適化がさらに必要であるが、ディープラーニングを用いた学習評定の推定において、小規模の学習データに対する一定の有効性が認められた。

キーワード：ディープラーニング, VBA, 評定, 学習評価, 観点別評価

## 1. 緒言

平成29年3月に告示された中学校学習指導要領総則<sup>1)</sup>では、学習評価について「多面的・多角的な評価を行っていくことが必要である。」と記述されている。さらに、「単元や題材など内容や時間のまとまりを見通しながら評価の場面や方法を工夫して、(以下略)」ともあり、単元や題材などのまとまりを3つの観点から各々評価することも求められている。このことから、ペーパーテストの点数だけではない学習評価の多様化が求められている。また、国立教育政策研究所が公表した学習評価<sup>2)</sup>は、「評定の適切な決定方法等については、各学校において定める。」と記述されている。このことから、教科担任制である中学校の場合、多様化した学習評価の各項目から総括的に評定することが明確に定まっておらず、教科の教員に任されることから評定に関して教員の裁量の余地が大きい。

一方、様々な分野でAI（人工知能）の技術が応用され、画像認識等において大きな成果を挙げている<sup>3)</sup>。教育分野においては、学習結果からどういった単元の学習が必要か推定するatama+<sup>4)</sup>と呼ぶAIが有名である。また、技術教育分野においても、このAIを学習内容に取り入れようとする試みもある<sup>5,6)</sup>。このようなAIの中でも特に大量のビッグデータを対象にしたディープラーニングと呼ばれる機械学習の1手法<sup>7)</sup>が様々な分野に適用され成果を挙げている。教員の過去の学習評価データからその特徴を学習させ、当該年度の学習評価から5段階の評定を推定させることは、このディープラーニングの対象となるタスクになり得ると当初考えた。しかしながら、ディープラーニングに与えることのできる学習データが十分確保できないことや、教員や教科、年度によって学習評価の項目数が著しく変化するという学習評価の特徴から、ディープラーニングには適用が難しいのではないかと懸念もあった。

学校における学習評価については、Excelの表計算機能を用いて集計等を行っている場合が多い。さらに

学習評価用の簡単なマクロ（VBA：Visual Basic for Applications<sup>8)</sup>）を有する Excel シートもフリーソフトとして提供されている<sup>9)</sup>。これを用いてディープラーニングを実現すれば Excel の中に評価推定機能を組み込むことができると考えた。Python 言語を用いたディープラーニングを参考にして、LiCLOG<sup>10)</sup> は VBA を用いて 3 層のニューラルネットワークを実現し、手書き数字の文字認識を可能にした。

本研究では、学習評価のデータに基づいた評価作業を行うために、今までの学習評価データと評価結果から、新たな学習評価に対して評価を推論するプログラムを試作した。これは、先の LiCLOG の手書き数字認識プログラムをベースにして、学習評価をニューラルネットワークの学習データとした評価プログラムへ改良した。さらに、これを拡張し、4 層のディープラーニングを実行するプログラムも併せて試作した。学習評価のデータとしては、大学学部の 2 科目（基礎情報学（H15～R3）、情報機器の操作（H17～R3））の成績データを対象とした。これらの前半の 8 割程度を学習データとし、残りを推論のテストデータに振り分けた。この試作した評価推定プログラムがどの設定であれば正しい推定を行うことができるのか、あるいは推定結果が利用可能なレベルにあるのか、2 つの学部データを事例として種々検討した。

## 2. ディープラーニングについて

人間の脳内には、千数百億個にもなるニューロン（神経細胞）があり、そのニューロンはお互いに電気信号を発して情報をやり取りしている。ニューロンは、長い軸索と木の枝のように分岐した樹状突起が伸びており、この樹状突起が、別のニューロンとつながって神経回路を形成している。この大量のニューロンが電気信号をリレーすることによって、何かを認識し、思考している（図 1 参照）。ニューラルネットワークは、図 1 に示すように人間のニューロンを模して作られ（これを人工ニューロンと呼ぶ）、これを複数つないでネットワークを構成する。

ニューラルネットワークは、図 2 のように 3 層（入力層、中間層、出力層）構造を持ち、1 つの層のすべての人工ニューロンが次の層の人工ニューロンに接続する構成になっている。電気信号をリレーする代わりに入力値に係数を乗し、各入力からの総和が閾値を超えたときに発火する。その係数が「学習」によって修正され最適化される。最適化された係数を用いて入力データに対して「推論」を行う。なお、ディープラーニングは図 2 のニューラルネットワークの中間層を 2 層以上に増やしたものを用いる。図中の○は、この人工ニューロンを示している。

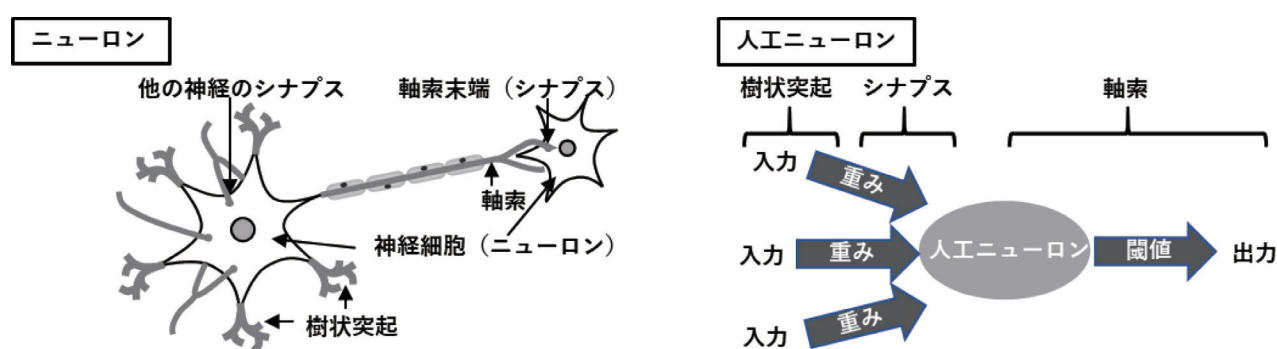


図1 ニューロンと人工ニューロン

ニューラルネットワークが学習として特徴を分類する場合、まず、あらかじめ特徴とそれに対する正答を含むデータを大量に読み込ませ、特徴から答えへと向かう計算のプロセスを学習させ、主に「重み」のパラメータとして記録しておく。

本研究では、図3に示すように成績データの各個人データ（シートの行）を入力データとして与え、重みとなる係数をそれぞれ掛けて中間層の値を求める。(1)式は、入力層の行列  $X$  と係数行列  $W$  の内積  $(\cdot)$  にバイアス行列  $B$  を加えて中間層の行列  $Y$  を求める。バイアス行列とは、各層に一定の値で底上げするための係数行列である。

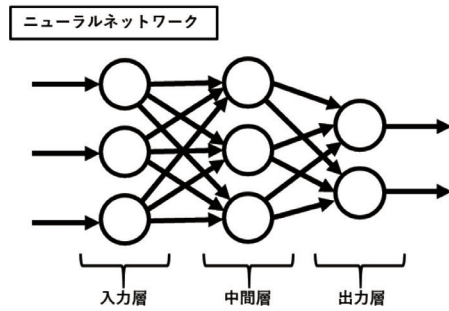


図2 ニューラルネットワークの模式図

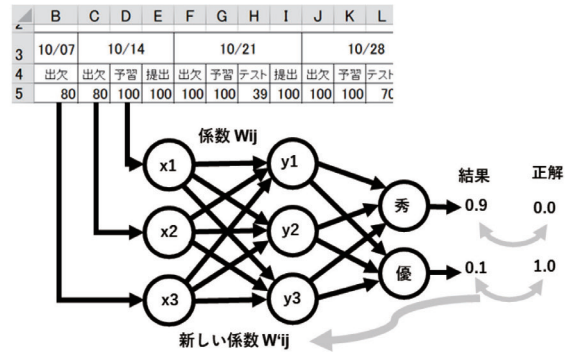


図3 ニューラルネットワークの学習

$$Y = W \cdot X + B \quad (1)$$

$$X' = \text{ReLU}(Y) \quad (2)$$

$$\text{ReLU}(x) = \begin{cases} x \geq 0 & x \\ x < 0 & 0 \end{cases} \quad (3)$$

(2) 式のように求められた行列  $Y$  からさらに、ReLU 関数を適用して次層（出力層）への入力値  $X'$  を求める。ReLU 関数とは入力  $x$  の値が 0 以上であれば、 $x$  を返し、0 未満であれば 0 を返す閾値関数である。次の層においても入力値  $X'$  から (1) および (2) 式を適用して出力  $Y'$  を求める。最後に、出力層として各評定の確率が Softmax 関数を用いて出力される。出力層  $Z$  の各要素  $z_k$  は、式 (4) に示される出力  $Y'$  の各要素  $y'_k$  から求められる。 $C$  は、指数関数のオーバーフローを防止するために挿入している定数である。

$$z_k = \frac{e^{(y'_k + C)}}{\sum_{i=1}^n e^{(y'_i + C)}} \quad (4)$$

すでに評定済みの正解と突き合わせ、例えば秀の結果が 0.9 であるが正解は 0（秀ではない）、優の結果から誤差をそれぞれ計算して、新しい係数行列  $W'$  とバイアス行列  $B'$  を算定する（誤差逆伝搬）。このプロセスを何度も繰り返すことで、正しい係数に収束していく。このプロセスを「学習」と呼ぶ。

$$W' = W - \alpha X \cdot Z^T \quad (5)$$

$$B' = B - \alpha Z^T \quad (6)$$

$$Z' = Z \cdot W^T \quad (7)$$

この誤差逆伝搬について簡単に説明する。(5) 式のように入力値  $X$  と Softmax 関数の出力  $Z$  を転置したものの内積に学習率  $\alpha$  を掛けた変化量を元の係数行列  $W$  から引いて学習した新たな係数行列  $W'$  を求める。学習率とは、変化の割合を設定するもので通常「0.001 ~ 0.01」という値が設定される。学習の精度によって変化させることができる。(6) 式は、バイアス行列  $B$  を同じく学習させ  $B'$  を求める。前の層への逆伝搬させるために誤差  $Z'$  を求め、これを前の層の出力  $Z$  とみなす。

ニューラルネットワークには、様々な関数がいわれているが、画像認識のタスクにおいて計算量が少なく比較的精度のよいものを採用した。

隠れ層の活性化関数：ReLU

出力層の活性化関数：Softmax

損失関数：交差エントロピー誤差

損失関数とは、推定して得られた結果と正解との誤りの度合いを示す関数であり、(8) 式に示すような交差エントロピー誤差関数を採用している。 $t_k$  は正解であり、 $y_k$  はニューラルネットワークの出力である。当然誤差  $E$  の値が小さい方が良い結果を示している。

$$E = -\sum_k t_k \log(y_k) \quad (8)$$

### 3. 試作したプログラム

Excel の book 自体の構成は学習フェーズで使用する学習用データが記入されている“train”と名付けたシート、学習が完了し、更新されたパラメータの精度を調べるためのテスト用データが記入された“test”と名付けたシート、そして決定したパラメータを保存し、再使用できるように書き込むための“parameters”という3枚のシートで構成している。VBA によって書いたプログラムはこのシート自体に適用しているわけではなく、標準モジュール、クラスモジュールとして保存している。標準モジュールとは基本的に VBA でプログラムを書く際に使うモジュールで、どのシートにも依存することがなく、すべてのシートから使用できる変数や、関数を宣言する際に使用する。クラスモジュールは、「同じ仲間の変数と関数をクラスとしてまとめる」ことができるため、全体のプログラムにおける変数名や関数名の競合を避けることができる。

図4は、この Main モジュールの処理概要をフローチャートで示したものである。なお、これは3層ニューラルネットワークのものである。まず、行列変数の初期化を行うが、係数行列には乱数を代入し、バイアス行列には0を代入する。続いて、学習データをシートから読み込み、0以上1未満の正規化を行う。続いてテストデータを読み込み、同様に正規化を行う。これを入力データとして行列演算を行い、 $Y$ を求め、さらに閾値関数 ReLU を適用して次層への入力行列を求める。出力行列  $D$  と正解の照合を行い、損失関数から誤差逆伝搬により係数行列およびバイアス行列を更新する。求めた各行列を適用して学習データおよびテストデータの正答率を求める。これを500回繰り返す。

表1 試作したプログラムのモジュール

分類	モジュール名	説明
標準モジュール	Main	メインモジュール 主プログラム
	function	汎用の関数群
クラスモジュール	Affine_layer	Affine 層の計算プログラム
	ReLU_layer	活性化関数
	SoftmaxWithLoss_layer	ソフトマックス関数損失関数
	TwoLayerNet	ニューラルネットワークの計算プログラム

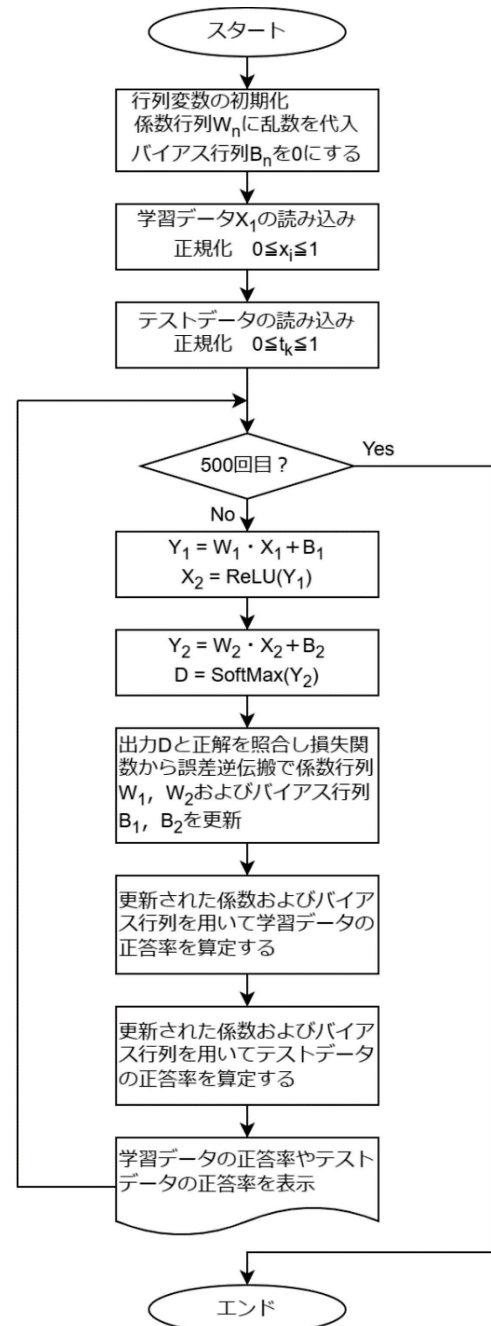


図4 Main モジュールの処理概要フロー



#### 4. 試行調査について

試行調査に適用した学習評価データは、学部授業である基礎情報学（H15年度～R3年度）と情報機器の操作（H17年度～R3年度）の成績データを学習用データとテスト用データに分け、学習用データによって学習した結果、得られたパラメータに基づいてテスト用データから「秀・優・良・可・不可」を推論する。各データは、受講者の講義日毎の出欠、小テスト結果、演習課題の提出状況、予習の提出状況、期末試験結果といった評価項目が縦の列毎に記載されている。

以降、基礎情報学の成績データを A データ、情報機器の操作を B データと記述する。図 5 にこの A データの成績シートを示す。なお、Excel の train シート中の A 列には評価結果（秀：4・優：3・良：2・可：1・不可：0）を入れ、期末試験結果については、項目の列の最後ではなく、B 列に設定している。C 列以降が本来の成績データとなる。また、マイナス点についてはすべて 0 点として訂正している。図 6 に、この変換された train シートを示す。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1																												
2																												
3	学籍番号	10/07	10/14		10/21		10/28		11/04		11/18		11/25		12/02													
4		出欠	出欠	予習	提出	出欠	予習	テスト	提出	出欠	予習	テスト	提出	出欠	予習	テスト	提出	出欠	予習	テスト	提出	出欠	テスト	提出	出欠	予習	テスト	提出
5		80	80	100	100	100	100	39	100	100	100	70	100	100	100	50	100	100	100	50	100	100	10	100	100	100	60	100
6		100	100	100	100	100	100	64	100	100	100	45	100	100	100	55	100	100	100	45	100	100	54	100	100	100	40	100
7		100	100	100	100	100	100	56	100	100	100	50	100	100	100	20	100	100	100	70	100	100	48	100	100	100	10	100
8		100	100	100	100	100	100	43	100	100	100	80	100	100	100	50	100	100	100	80	100	100	72	100	100	100	40	100
9		80	80	100	100	100	100	37	100	100	100	80	100	100	100	35	100	100	100	20	100	100	50	100	100	100	40	100
10		80	80	100	100	100	100	64	100	100	100	70	100	100	100	55	100	100	100	80	100	100	34	100	100	100	80	100

図 5 基礎情報学の成績シート

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
44	2	40	80	80	100	100	100	100	39	100	100	100	70	100	100	100	
45	4	82	100	100	100	100	100	100	64	100	100	100	45	100	100	100	
46	3	73	100	100	100	100	100	100	56	100	100	100	50	100	100	100	
47	4	100	100	100	100	100	100	100	43	100	100	100	80	100	100	100	
48	4	75	80	80	100	100	100	100	37	100	100	100	80	100	100	100	
49	4	100	80	80	100	100	100	100	64	100	100	100	70	100	100	100	

図 6 基礎情報学の train シート

図 7 のように A データの学習用データとして使用するのは平成 15 年から平成 30 年までの 15 年分のデータで、評価項目数は決まっておらず年度毎にばらけている。テスト用データは令和元年から令和 3 年のデータを使用している。同じく B データの年度別項目数を図 8 に示す。

また、表 2 の標準偏差のように A データと比べて項目数が比較的一定であるため、比較用データとして学部授業の情報機器の操作の成績データも調査に使用する。

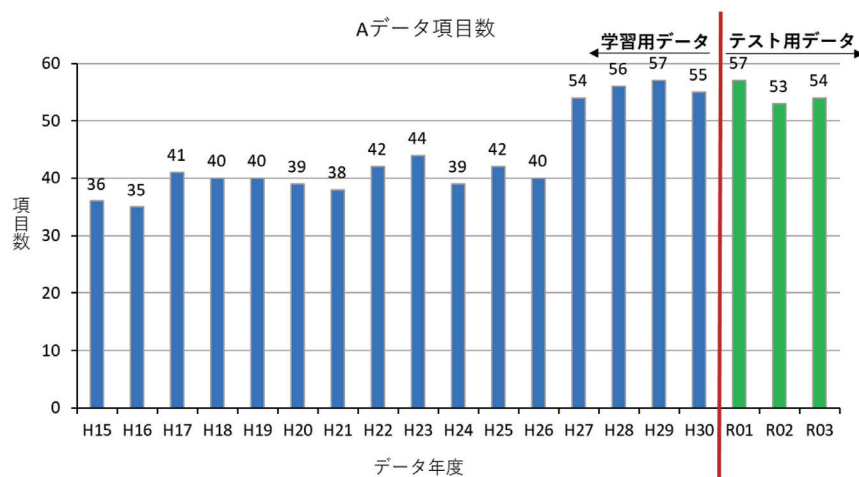


図 7 A データの年度別項目数

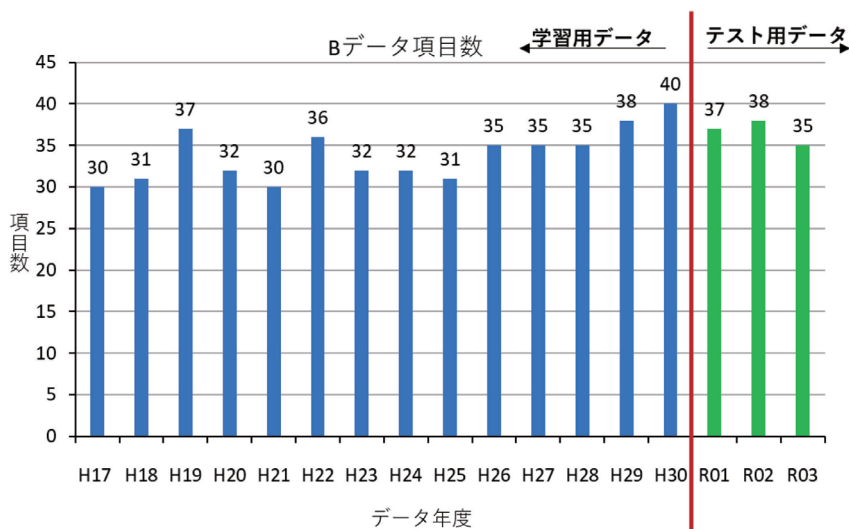


図8 Bデータの年度別項目数

表2 両データの項目数の平均と標準偏差

項目	Aデータ	Bデータ
平均	45.37	34.35
標準偏差	7.76	3.01

今回試作した推定プログラムが実用的な精度を有するものなのか、また、ニューラルネットワークには層数や各層のニューロン数等の設定項目が多数あることから、最適な設定方法を探る必要もある。そのために推定プログラムに前節の学習評価データを与えて、その推論フェーズにおけるテスト用データの正答率を調査することで最適な設定条件を探り、どの程度まで正答率を上げることができるのか見極める。調査方法はAデータとBデータともに最初のパラメータが乱数であるため、学習の平均をとるために、学習と推論を3回行い、テスト用データの最大正答率、その時の学習回数の平均をとる。

## 5. 試行調査結果

ニューラルネットワーク（3層）、ディープラーニング（4層）のそれぞれをAデータの学習用データのテスト用データの正答率を比較する。また、その際の中間層の数とテスト用データ正答率の関係も棒グラフ（図9参照）で表している。

また、表3にニューラルネットワークとディープラーニングの試行結果を示す。図10には500回の学習における学習データの正答率およびテストデータの正答率、加えて損失関数の値の推移を示す。図中の赤丸には、最大値に達した後にテストデータの正答率が低下していることから過学習の兆候も認められる。

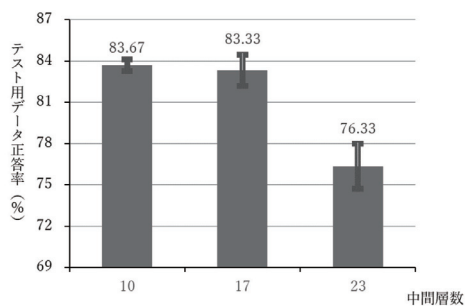


図9 Aデータの4層の中間層1の層数と正答率

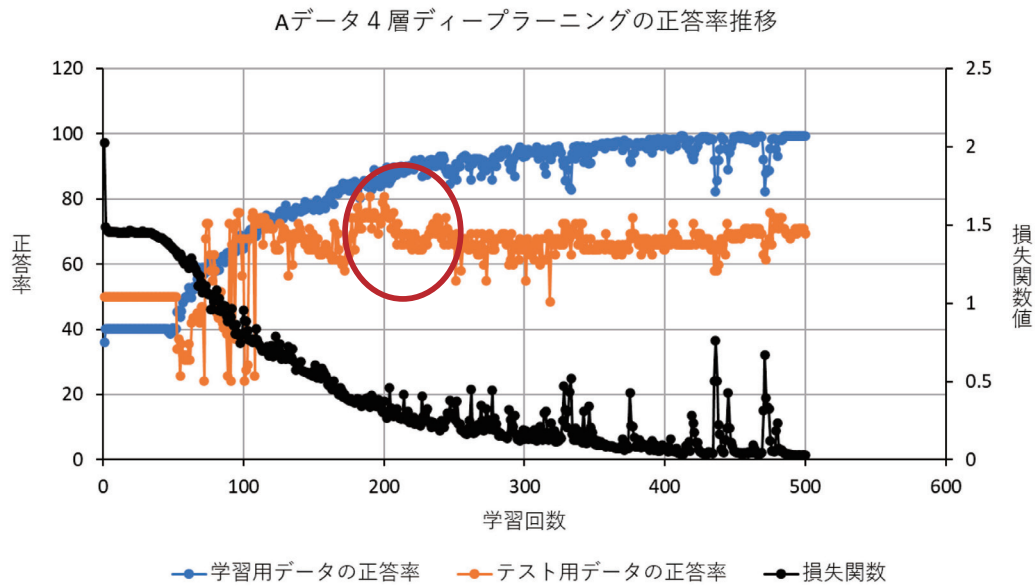


図10 Aデータの4層の正答率等の推移

表3 Aデータ試行結果まとめ

3層		4層	
最大正答率	76.67%	最大正答率	83.67%
第二候補	89.33%	第二候補	94.00%
中間層数	17	中間層数	32   10
学習回数	128	学習回数	113

表4 Bデータ試行結果まとめ

3層		4層	
最大正答率	72.67%	最大正答率	77.33%
第二候補	84.33%	第二候補	82.00%
中間層数	17	中間層数	32   17
学習回数	353.33	学習回数	244.00

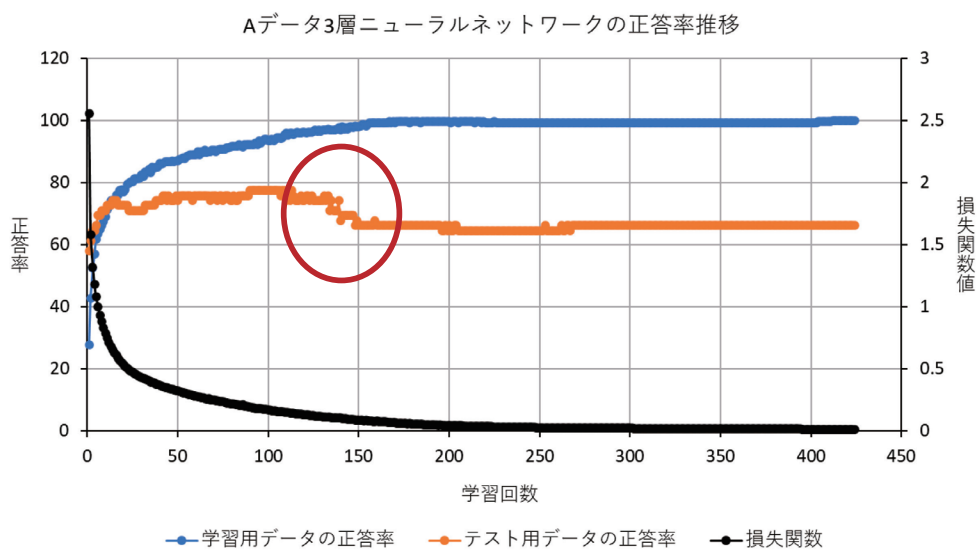


図11 3層ニューラルネットワークのAデータでの過学習

表4には、Bデータの試行結果のまとめを示す。正答率がAデータに比べて低下している。また、第二候補までを正答とした場合の正答率については、表3に示すように94.0%という値が得られた。また、データ数が少ないことで過学習を起こしていることも予想される。過学習とは、学習用データから学習しすぎた結果、学習データに特化したパラメータになってしまい、テスト用データの推論がうまくいかなくなってしまうことである。すなわち、学習フェーズによって得た推論のプロセスに汎用性がなくなっている状態のことである。図10および図11を見ると、テスト用データのテスト用データの正答率が最大になったところから低下していることがわかる。これは過学習によって、テスト用データの正答率が上がっていないという予想が立てられる。データ数が少ないと学習できる情報も減り、データの学習が終わるのが早い。だが、学習用データのテスト用データの正答率も上げなければならず、続けて学習させるため、過学習に陥りやすいことも認められた。

## 6. 結言

本研究では、学校での利用を想定したニューラルネットワークおよびディープラーニングを用いた成績評定推定プログラムを試作し、試行させながらプログラムに適用する学習用データや、中間層の数などを比較検討した。ディープラーニング（4層）であれば第二候補までを含めた正答率から学部科目での利用が可能であるレベルにあることが認められた。また、少ない学習用データから過学習の兆候も認められた。

今後の課題としては、テスト用データの正答率をさらに上げることや、4層ディープラーニングから、5層、6層と増やした試行も検討する必要がある。3層よりも4層の方が正答率の向上が認められたことから、層数をさらに増やすことも検討してみる必要がある。また、中学校現場における学習評価データから評定を推定することも検討する必要がある。

## 参考文献

- 1) 文部科学省：中学校学習指導要領（平成29年告示）解説総則編，（2019）
- 2) 国立教育政策研究所：「指導と評価の一体化」のための学習評価に関する参考資料 中学校 技術・家庭，（2020）
- 3) AI白書編集委員会：AI白書2022，KADOKAWA，（2022）
- 4) EdTech研究所：「公式」AI教材 atama+（アタマプラス），<https://www.atama.plus/>
- 5) 高橋典弘：AIに対する中学生の意識調査とカリキュラムモデルの提案，日本産業技術教育学会九州支部論文集，第29巻，pp.11-17，（2021）
- 6) 高橋典弘，梅野貴俊，下戸健：機械学習の基本が理解できるプログラミング学習教材の開発，日本産業技術教育学会誌，第64巻，第2号，pp.103-110，（2022）
- 7) 斎藤康毅：ゼロから作るDeep Learning —Pythonで学ぶディープラーニングの理論と実装，オライリージャパン，（2016）
- 8) 井上豊：改訂版 教師のためのExcelVBA活用法 Excelを200%活用する！初級者のステップアップ読本，明治図書出版，（2016）
- 9) 佐野嘉昭：エクセルファイルの紹介，<http://www.kyoinexcel.com/file.html>
- 10) LiCLOG：【DeepLearning】Excel VBAでニューラルネットワークを再現 | MNIST学習で文字認識，<https://liclog.net/excel-vba-deeplearning-mnist/>