

深層学習による投入係数行列の予測

福井 昭吾

概要

本研究では、深層学習により産業連関表の投入係数行列を推計する新たな手法を提示する。この研究では都道府県のデータを用いた深層学習を試みるが、そのデータが小さいために過学習を引き起こす可能性が高い。そこで、複数の都道府県を一つにまとめた仮想的な地域のデータを生成するデータ拡張により、データサイズを増加させ過学習の回避を試みる。山口県および岐阜県郡上市について投入係数行列を予測した結果、深層学習による推計方法はRAS法より安定した精度で投入係数を予測できうることを示した。

キーワード 深層学習、産業連関表、投入係数

Abstract

In this study, we present a new method for estimating input coefficient matrix with deep learning. We use the data of prefectures of Japan as the training data. However, the data is small and is likely to cause over-fitting. Therefore, we attempt to avoid the over-fitting by creating new data of virtual areas aggregating multiple prefectures. The approach is based on the concept of data augmentation. As a result of predicting the input coefficient matrix of Yamaguchi prefecture and Gujo city, we showed that the method using deep learning can estimate the matrix with more stable accuracy than RAS method for these areas.

Keywords Deep Learning, Input-Output Table, Input Coefficient

1 はじめに

産業連関表は、マクロ的視点で経済分析を行う際の重要な基礎データである。例えば、政策やイベントの経済効果を分析する際には、産業連関表からレオンチェフ逆行列を導出し経済効果の金額を試算することが多い。また、応用一般均衡分析で用いる社会会計表は産業連関表に基づいて作成される。

産業連関表の作成には多くの一次統計が必要であり、かつ、それら一次統計に基づく膨大な計算が必要となる。そのため、公表される産業連関表の地域・年は極めて限定されている。2021年時点では、一次統計に基づいて作成された産業連関表は日本全体を対象とするもののみで、都道府県や市町村の産業連関表は、追加的な調査と何らかの推計手法を組み合わせたハイブリッド法により推計されている。さらに、市町村レベルでは産業連関表の推計に必要な統計が大きく不足するため、産業連関表を公表している市町村は極めて少ない。また、「国勢調査」や「経済センサス - 基礎調査」など、産業連関表の作成で特に重要な役割を果たす一次統計は複数年の間隔を置いて調査・公表が行われていることから、産業連関表は5年おきのみ公表されている。

産業連関表が公開されていない地域・年について産業連関表が必要な場合、何らかの方法で産業連関表を推計することになる。産業連関表の推計では、中間投入および投入係数の推計に大きな困難を伴う。これまで、特に投入係数を推測するための様々な手法が開発されてきた。それらの手法の中で、現在はRAS法と呼ばれる手法が広く用いられている。RAS法は、初期時点における投入係数行列、および、対象時点の各産業の生産額・中間需要・中間投入をもとに、対象時点の投入係数行列を推計する手法である。McDougall (1999)によれば、イギリスの産業連関表の構築に際してR.StoneがRAS法を使用したことが、RAS法が幅広く利用されるようになった端緒とされる。RAS法では、初期時点の投入係数行列と対象時点の生産額に基づいて中間投入行列の推計値を求め、中間投入行列の行和と対象時点における各産業の中間需要が

等しくなるように、かつ、中間投入行列の列和と対象時点における各産業の中間投入が等しくなるように、投入係数の値の更新を繰り返すことで投入係数行列を求める。本来のRAS法は、過去時点における既知の投入係数行列に基づいて将来時点の投入係数行列を求める。しかし、日吉他 (2004) が試みているように、時間軸を空間軸に置き換えることで地域の産業連関表についても投入係数行列の推計が可能となる。

RAS法は推計に必要な計算が分かりやすいことから、投入係数を推計する際に広く利用されている。一方、RAS法による推計には経済学的根拠が薄いという指摘がある。RAS法の過程では、加工度変化修正係数と代替変化修正係数を計算して投入係数を更新していくが、岩崎 (1980) は、これら係数と現実の技術変化との間の経済的意味付けが把握されていないこと、加工度の変化と代替の変化を同一率で行うことは極端な仮定であることを指摘している。またRAS法では、投入係数の推計値が、対象時点の産業ごとの生産額・中間需要・中間投入の値に依存する性質がある。例えば、長谷川 (1991) では産業ごとの生産額・中間需要・中間投入に対して誤差による変動を加えることで、RAS法により推計される投入係数行列の誤差についてシミュレーションを行っている。その結果、生産額・中間需要・中間投入における誤差は、投入係数の推計値に一定程度の偏りをもたらすことを示している。

本研究では、地域を対象とする産業連関表に着目し、RAS法の代替となりうる投入係数行列の新たな推計方法を考察する。この新しい推計方法は、深層学習により投入係数行列を予測するものである。

ニューラルネットワークモデルは、説明変数と目的変数との非線形関係をモデル化するもので、目的変数について高い精度での予測を行うために導入される。McCulloch and Pitts (1943) の研究をニューラルネットワークモデルの端緒として、モデルの学習手法の改良や計算機の性能向上を経て深層学習 (deep learning) が生み出された。ニューラルネットワークモデルを構成する活性化関数の層を多層化したものを深層ニューラルネットワークモデルという。深層学習とは、深層ニューラルネットワークモデルを用いて説明変数と

目的変数との間の複雑な関係を精緻にとらえ、目的変数について高精度の予測を試みるものである。

投入係数の推測にニューラルネットワークモデルを用いた研究として、Papadas and Hutchinson (2002)がある。彼らの研究では、産業ごとの中間投入と生産額の比・産業ごとの中間需要と生産額の比・産業ダミーを説明変数、投入係数を目的変数として、それらの間にニューラルネットワークモデルを仮定して投入係数の予測を試みている。しかし、イギリスにおける1984年の7部門産業連関表の投入係数を目的変数のデータとして用いているため、データサイズは $7 \times 7 = 49$ と非常に小さい。また、予測に用いているニューラルネットワークモデルは、三つの層（入力層・一つの間層・出力層）のみから構成され、間層のノードも7個とそれほど複雑なモデルではない。多層・多ノードなニューラルネットワークモデルを小さなデータで学習した場合、モデルがデータに過剰に適合し予測精度が低下する過学習が発生することから、49というデータサイズでは複雑なモデルの構築が困難だったと推察する。さらに、投入係数は比較的高い精度で予測できているものの、RAS法よりも予測の精度は低い。

学習に用いるデータサイズの問題に対処するために、本研究ではデータ拡張の適用可能性について考える。機械学習の分野では、本来のデータに対してデータを増加させてモデル学習を行うデータ拡張という方法が使われている。データ拡張は、モデルの予測精度の向上をもたらすとされる。例えば、Bishop (1995) は不変性を持つ訓練データについて、入力にノイズを与えることによるデータ拡張は、一般化Tikhonov正則化型の項を誤差関数に加えることと等価であることを示した。誤差関数に正則化項を加えて最適化を行うことで、過学習が抑制されることが分かっている。そのため、入力にノイズを与える形のデータ拡張は、過学習を抑え汎化性能を向上させる。また、Wu et al. (2020)は、画像データに対する各種のデータ拡張が汎化性能を向上させることを、実際のモデル学習により示している。

Papadas and Hutchinson (2002)が発表された2002年と比較して、2021年現

在では計算機の性能が大幅に向上しているため、深層ニューラルネットワークモデルを現実的な計算時間で学習できるようになった。また、上述のデータ拡張を用いることで、データサイズの不足に伴う過学習に対処できる可能性がある。本研究では、複数の地域を仮想的な一地域にまとめるデータ拡張を導入することで、深層学習による高い精度での投入係数の予測を試みる。

2 ニューラルネットワークモデルとは

ニューラルネットワークモデルとは、活性化関数と呼ばれる関数を並列し層にして結合することで、モデルの入力である説明変数と出力である目的変数との間の非線形関係を表現するモデルである。一般に、ニューラルネットワークモデルは目的変数を高い精度で予測するために利用される。ニューラルネットワークを構成する関数を活性化関数という。活性化関数としては、シグモイド関数やReLU(rectified linear unit)関数など、簡単な形の非線形関数を設定することが多い。深層ニューラルネットワークモデル（以下、DNNモデルと呼ぶ）は、ニューラルネットワークモデルの層を多層化したもので、DNNモデルを用いた学習を深層学習という。その予測精度の高さから、近年では画像認識・文章解析・行動予測といった分野で深層学習が積極的に用いられている。

図1にはニューラルネットワークモデルの例を示している。このニューラルネットワークモデルは、二つの中間層と一つの出力層からなる。1番目の中間層は M_1 個のノード、2番目の中間層は M_2 個のノードで構成され、各ノードに活性化関数が配置される。出力層では出力の数に等しいノードを置く。例えばニューラルネットワークモデルからの出力が1個の場合、出力ユニット活性と呼ばれる1個の関数を出力層のノードに設定する。図1のモデルについて入力から出力までの流れを見てみよう。始めに、説明変数ベクトル $\mathbf{x} = (x_1, x_2, \dots, x_K)'$ が第1の中間層に入力され M_1 個の出力が得られる。次に、これらの M_1 個の出力が第2の中間層の入力となり、その結果 M_2 個の出力

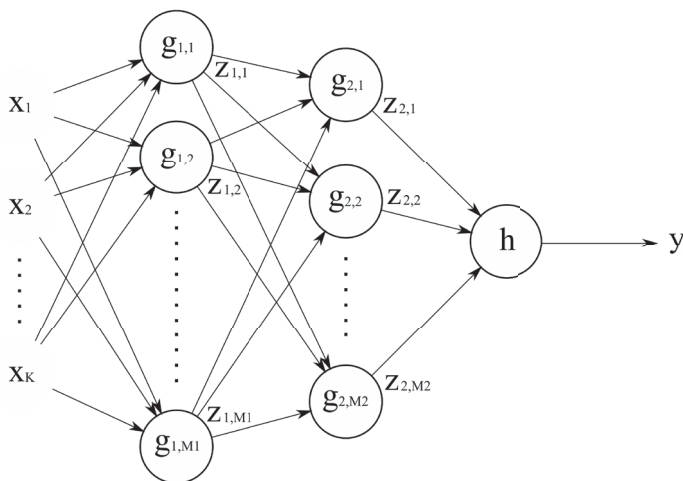


図 1: 二つの中間層と一つの出力層からなるニューラルネットワークの例

が得られる。最後に、これら M_2 個の出力が出力層に対する入力になることで目的変数 y が求められる。ニューラルネットワークモデルでは、層の数や大きさを変えることで複雑な非線形性を表現でき、また活性化関数ごとに入力変数を変えるなどの形で層の構造を柔軟に拡張することも可能である。

ニューラルネットワークモデルによる予測を行うためには、モデルの学習が必要である。モデルの学習とは、ニューラルネットワークモデルに含まれる活性化関数等のパラメータをデータに基づいて推定することである。図1のニューラルネットワークモデルで、 i 番目の中間層 ($i = 1, 2$) における第 j ノードの活性化関数 $g_{i,j}$ を

$$z_{i,j} = g_{i,j}(w'_{ij} z_{i-1})$$

と表す。ただし、 w_{ij} はパラメータ、 z_{i-1} は活性化関数の入力変数、 $z_{i,j}$ は活性化関数の出力変数である。なお、 $z_0 = x$ とする。また、出力ユニット活性を

$$y = h(w'_o z_2)$$

とする。ただし、 $\mathbf{w}_o$ はパラメータベクトル、 $\mathbf{z}_2$ は2番目の中間層からの出力値からなるベクトルで $\mathbf{z}_2 = (z_{2,1}, \dots, z_{2,M_2})'$ である。例えば目的変数が連続量を取る場合は、出力ユニット活性として恒等写像 $\mathbf{w}'_o \mathbf{z}_2$ を用いる。なお、パラメータベクトル $\mathbf{w}_{ij}$ および $\mathbf{w}_o$ は定数項を含んでいる。説明変数の数を K 個とすると、1番目の層における各活性化関数のパラメータはそれぞれ $K+1$ 個、2番目の層における各活性化関数のパラメータは M_1+1 個、出力ユニット活性のパラメータは M_2+1 個となり、モデル全体としては $(K+1)M_1 + (M_1+1)M_2 + (M_2+1)$ 個のパラメータを含む。

モデルの学習、すなわち活性化関数のパラメータの推定は誤差関数を最小化し実現する。いま、すべてのパラメータからなるベクトルを $\mathbf{w}$ とし、上記のニューラルネットワークモデルを $y(\mathbf{x}; \mathbf{w})$ と表す。説明変数および目的変数のデータの組 $\{(\mathbf{x}_i^*, y_i^*) | i = 1, \dots, N\}$ が得られたとき、誤差関数

$$E(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (y_i^* - y(\mathbf{x}_i^*; \mathbf{w}))^2$$

を最小化する $\mathbf{w}$ の値をその推定値とする。実際には、計量経済分析における最尤法による数値解の導出と同様、この誤差関数のパラメータについての偏微分を求め、数値最適化を使って誤差関数の値が最小になるようなパラメータの値を探索することで推定値を求める^{*1}。

ニューラルネットワークモデルにおけるパラメータの偏微分は、ネットワーク構造を逆にたどることで解析的に導出する。この手法はRumelhart et al. (1986)によって確立され、逆誤差伝播法と呼ばれる。例えば上述のニューラルネットワークモデルで $y_i = y(\mathbf{x}_i; \mathbf{w})$ として、 $\mathbf{w}_{11}$ の最初のパラ

^{*1} 実際には、誤差関数はパラメータについて非線形であることが多いため、数値最適化により誤差関数の大局的極小値が求められるとは限らない。

メータ $w_{11,0}$ について誤差関数に対する偏微分を求めると、連鎖公式により

$$\begin{aligned}\frac{\partial E}{\partial w_{11,0}} &= \sum_{i=1}^N \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial w_{11,0}} = \sum_{i=1}^N \frac{\partial E}{\partial y_i} \sum_{k=1}^{M_2} \frac{\partial y_i}{\partial z_{2,k}} \frac{\partial z_{2,k}}{\partial w_{11,0}} \\ &= \sum_{i=1}^N \frac{\partial E}{\partial y_i} \sum_{k=1}^{M_2} \frac{\partial y_i}{\partial z_{2,k}} \sum_{l=1}^{M_1} \frac{\partial z_{2,k}}{\partial z_{1,l}} \frac{\partial z_{1,l}}{\partial w_{11,0}}\end{aligned}$$

となる。これより、活性化関数・出力ユニット活性・誤差関数について解析的に偏微分を求めることができるならば、誤差関数のパラメータについての偏微分 $\partial E / \partial w_{11,0}$ もまた解析的に求めることが可能となる。

以上の偏微分を用いて、数値最適化によりパラメータの推定値を求める。ニューラルネットワークモデルの学習では、勾配降下法を土台とする最適化法を用いることが多い。勾配降下法は、最適化の対象となる目的関数についてパラメータの偏微分ベクトルを求めパラメータの値を負の偏微分の方角にわずかに更新するという処理を逐次的に行うことで、数値的に極小値を見つける方法である。実際の学習では、確率的勾配降下法やAdamといった方法を用いる。

これまでの説明から明らかであるように、ニューラルネットワークモデルの学習と計量経済分析におけるモデル推定は、理論上ほぼ同じ手続きで行われる。一般的な計量経済分析では、線形モデルや多項選択モデルなどの非線形モデルを説明変数と目的変数との間に仮定し、残差平方和や尤度関数といった目的関数についてパラメータによる偏微分をもとに最適化を行い推定値を求める。ニューラルネットワークモデルの学習は、計量経済分析のモデル推定の手続きにおいて、線形モデルや多項選択モデルがニューラルネットワークモデルに置き換わるに過ぎない。ただし、実際の最適化に大きな違いがある。計量経済分析におけるモデル推定の目的は、通常は、現状の目的変数を最も良く説明できるモデルの構築である。したがって、手元にあるすべてのデータを用いて最適化を行うことで推定値を導出しようとする。一方、ニューラルネットワークモデルを学習する目的は、目的変数を最も良く予測できるモデルを構築することである。つまり、学習に用いていない未知の

需要側 供給側		中間需要			中間 需要 計	最終 需要	移 出	移 入	輸 出	輸 入	生産額
		産業1	産業2	産業3							
中 間 投 入	産業1	m_{11}	m_{12}	m_{13}	$\sum_j m_{1j}$	F_1	L_1	N_1	E_1	I_1	Y_1
	産業2	m_{21}	m_{22}	m_{23}	$\sum_j m_{2j}$	F_2	L_2	N_2	E_2	I_2	Y_2
	産業3	m_{31}	m_{32}	m_{33}	$\sum_j m_{3j}$	F_3	L_3	N_3	E_3	I_3	Y_3
中間投入計		$\sum_i m_{i1}$	$\sum_i m_{i2}$	$\sum_i m_{i3}$							
付加価値 関連項目		V_1	V_2	V_3							
生産額		Y_1	Y_2	Y_3							

図 2: 3 産業からなる産業連関表

データについても高い予測精度（汎化性能）を持つことを期待する。すべてのデータを使って最適化を行う場合、学習したモデルはそのデータの説明に特化し未知のデータについての予測精度は低くなる傾向がある。そこで、取得したデータを訓練用データとテスト用データに分割し、訓練用データによりモデルを学習した後でテスト用データによる予測精度を求め、学習したモデルの汎化性能を判定するといった方法で、ニューラルネットワークモデルにおける最適化が行われる。

ニューラルネットワークモデルの学習が完了した後、そのモデルを使って推論（予測）を行う。推論したモデルに予測対象の説明変数の値を入力することで、目的変数の予測値を得ることができる。

3 深層学習による投入係数の予測

産業連関表は、対象とする地域で一定期間内に発生した生産物の中間投入・産出構造および最終需要などを産業部門ごとに記録し、金額表示にして

表形式にまとめたものである。図2は産業連関表の一例で、三つの産業部門からなる産業連関表の形状を示している。産業連関表の各行は各産業が生産した財に対する需要の構成を表し、各列は各産業の費用の構成を表している。この表のうち m_{ij} を中間投入といい、第 i 産業が第 j 産業に販売した中間生産物（原材料等）の金額、あるいは、第 j 産業が第 i 産業から購入した中間生産物の金額を表している。中間投入を第 (i, j) 要素とする行列を中間投入行列という。図2の産業連関表では、中間投入および最終需要は対象地域内の生産物の購入額に加えて対象地域外からの生産物の購入額も含んでいる。この場合、対象地域外からの生産物の購入額は、移入・輸入として差し引くことで相殺する。移入・輸入についてこのような形で取り扱う産業連関表を競争移輸入型産業連関表という^{*2}。投入係数は、生産物1単位の生産に必要な各産業の中間投入の額である。第 j 産業部門の生産量を Y_j と表すと、第 j 産業部門が1単位の生産を行うために必要とする第 i 産業からの中間投入の購入額は $a_{ij} = m_{ij}/Y_j$ と計算できる。この a_{ij} を投入係数という。また、投入係数 a_{ij} を第 (i, j) 要素として持つ行列 A を投入係数行列という。

産業連関表は、経済効果の算出や応用一般均衡分析といった様々な分析で用いられており、マクロ的視点における定量的な経済分析において重要な役割を果たすものである。しかし現状では、市町村ごとの産業連関表はほとんど公表されていないため、市町村を対象に産業連関表を用いた分析を行うことは極めて困難である。産業連関表の導出には各種の一次統計が必要となるが、それら一次統計の多くについて市町村別の集計結果が公表されていない。そのため、一次統計に基づいて市町村別の産業連関表を導出することが難しい。そこで、市町村を対象に産業連関表を用いた分析を行う場合は、何らかの方法に基づいて市町村別の産業連関表を推測する必要がある。

^{*2} 本研究では移輸入は外生的であるとみなして議論を進める。移輸入が内生的であるとする場合、移輸入を推測するための別の手法を考える必要がある。移輸入を内生的とみなして推測する手法としては Location Quotient(LQ) 法などがある。LQ 法の詳細については、例えば朝日 (2004) が詳しい。

産業連関表を推測する際、生産量・中間投入・最終需要など産業連関表を構成する部分ごとに推測が行われる。特に、中間投入または投入係数の推測については多くの議論があり、様々な推測方法が提示されてきた。現在、中間投入の推測では、RAS法により投入係数を推測した後で中間投入を計算するという方法が広く利用されている。RAS法とは、対象時点における産業ごとの生産額・中間投入・中間需要を用いて、初期時点の投入係数行列を更新することで対象時点の投入係数行列を求める手法である。はじめに、初期時点の投入係数行列 A_0 と対象時点の各産業の生産額に基づいて、各産業の一時的な中間投入と中間需要を求める。次に、対象時点の中間投入と一時的な中間投入の要素ごとの比（修正係数と呼ぶ）を求め、それらの比を対角要素とする行列 R を作成する。中間需要についても同様に行列 S を作る。その後、 $A_1 = RA_0S$ として投入係数行列を更新する。修正係数が十分に1に近づくまで上記の手順を繰り返すことで、投入係数の推計値を得る。

RAS法による投入係数の推測では、初期値となる投入係数行列、および、対象時点における産業ごとの生産額・中間投入・中間需要の値を設定する必要がある。長谷川 (1991) が示すように、RAS法の精度はこれらの値に影響を受ける。一般に、対象時点における各産業の生産額・中間投入・中間需要の正確な値を事前に知ることは難しいため、それらの予測値を求めることになる。したがって、RAS法を使って投入係数を高い精度で予測するならば、産業ごとの生産額・中間投入・中間需要の値についても高い精度で予測しなければならない。また、RAS法による投入係数行列の推測は、投入係数に影響する要因やその生成プロセスを考慮するものではなく、中間投入の行和および列和について予測値と目標値が近くなるような投入係数を推測値として選んでいるに過ぎない。

RAS法の持つこれらの問題点を解決しうる新たな手法として、以下では、深層学習により投入係数を予測する方法の提示を試みる。本研究で提示する手法は、投入係数を目的変数としてその生成過程をニューラルネットワークモデルにより説明することから、投入係数の要因とその生成プロセスを考慮

する方法といえる。また、深層学習に基づく方法では、投入係数を予測するために産業ごとの生産額・中間投入・中間需要を使うことは要求されないため、それらの値を事前に予測しておく必要がない。実際に後で示す実証分析では、産業ごとの生産額・中間投入・中間需要を使うことなく、深層学習により投入係数行列を予測している。

深層学習による投入係数の予測に際して、以下の仮定を置く。

仮定 1 すべての地域において、投入係数 a_{ij} の生成過程は以下の式 f_{ij} によって近似される。

$$a_{ij} = f_{ij}(x_1, x_2, \dots, x_K). \quad (1)$$

ただし、 $x_1, x_2, \dots, x_K$ は投入係数の説明変数である。

一国全体の産業連関表は一次統計に基づき定められた手順で推計されている。この手順が正しいとすれば、一次統計の値に対して投入係数の値は一意に決まる。これより、経済の状態が全く同じ二つの地域は、一次統計に含まれる各値・ f_{ij} で用いられる説明変数の値はすべて等しいため、投入係数もまた互いに同値になると仮定できる。この本来の投入係数の生成手順が f_{ij} によって近似されるならば、説明変数と投入係数間の一意性は f_{ij} においても維持されなければならない。上記の仮定は、この一意性を保持するためのものである。

この仮定により、投入係数の生成過程は、一次統計に基づく厳密な生成過程に対して近似による誤差と推計方法の違いに伴う誤差を加えたものとみなされる。モデルの学習に都道府県のデータを用いる場合、上記の仮定を置くことで、すべての都道府県で同一の方法で投入係数を計算すると仮定することになる。しかし実際には、都道府県レベルの産業連関表についてその作成に必要な一次統計が一部公表されず、各都道府県はハイブリッド法などの手法を使って投入係数を推計している。その際、投入係数の推計方法、すなわち生成過程 f_{ij} がすべての都道府県で同じであることは保証されない。それでも上記の仮定を置くことは、各都道府県の投入係数の生成過程が本来の生成

過程に近似誤差と推計方法の違いによる誤差を加えたものとみなすことを意味する。

以下では、 f_{ij} に対して深層ニューラルネットワークモデルを仮定し、都道府県のデータを用いてモデル学習を行うことを考える。説明変数 $x_1, x_2, \dots, x_K$ は、目的変数である投入係数の説明要因であることから、説明変数のデータとして中間投入および生産額の推計に用いる一次統計の値を用いることが望ましい。しかし上述の通り、都道府県について産業連関表の計算に必要な一次統計は一部公表されない。そこで本研究では、これら一次統計の値の代替として、生産性を表す変数を説明変数とする。

都道府県のデータを用いて f_{ij} についての深層学習を行う場合、データサイズが小さいことによる予測精度の低下が問題となる。小さいデータで深層学習を行った場合、学習に用いたデータに対してモデルが過剰に適合する過学習が起ころう。過学習が生じたモデルは未知のデータに対する当てはまりが低いためモデルの予測精度は低下する。そのため、深層学習では数千から数万のサイズのデータが必要であるとされる。いま、2015年時点の都道府県のデータに基づいてモデルを学習するとしよう。2021年2月時点で2015年産業連関表が公表されている地域は、日本全体、34道府県、および、極めて少数の市のみである。34道府県のデータに基づいて f_{ij} を学習するならば、標本の大きさは34となり深層学習を行うには極めて小さい。したがって、都道府県のデータのみを用いて投入係数についての深層学習を行う場合、標本が小さいために高い確率で過学習が発生すると予想する。

小さいデータを用いて深層学習を行う際、データ拡張という手法によりデータサイズを増加させて学習を試みることがある。データ拡張は主に画像認識で用いられる手法で、画像データの持つ不変性に基づいている。画像認識では、画像情報を説明変数、その画像が何か（ラベルという）を目的変数としてモデル学習を行う。通常、画像に対してわずかに回転・縮小などの加工を行ってもラベルは変わらない。この性質を画像データの不変性という。画像データの不変性を利用することで、元となる画像に回転・縮小などの加工

を施した画像情報を新たな説明変数の値、元の画像のラベルを新しい目的変数の値として、新しい観測値を生成することが可能となる。このようにしてデータを増やす手法をデータ拡張と呼ぶ。

Zhang et al. (2018)は、データ拡張の一手法としてmixupという方法を提示した。また、mixupにより予測精度が向上することを実証分析により示している。このmixupでは、訓練データ内の二つの観測値 (x_1, y_1) と (x_2, y_2) について、ベータ分布 $B(\alpha, \alpha)$ 、 $\alpha > 0$ より抽出した乱数 λ を使い、

$$\begin{aligned}\hat{x} &= \lambda x_1 + (1 - \lambda)x_2 \\ \hat{y} &= \lambda y_1 + (1 - \lambda)y_2\end{aligned}$$

として仮想的な観測値 $(\hat{x}, \hat{y})$ を生成する^{*3}。

以下では、mixupと同様に、任意の2地域（それぞれ $R1, R2$ と表す）をまとめた一つの地域 $R12$ を生成することによるデータ拡張の可能性を検討してみよう。データ拡張を行うためには、加工した説明変数の値に対して正しい目的変数の値が得られる必要がある。画像データの変換については、変換前と変換後とで目的変数の値は変わらないことが不変性より保証される。mixupの場合は、二つの観測値 (x_1, y_1) と (x_2, y_2) に対して、確率 λ で y_1 、確率 $(1 - \lambda)$ で y_2 と判定されるように画像データ x_1 と x_2 を合成したとみなせる。一般に、mixupは画像データや音声データなど合成可能なデータにのみ適用可能であり、目的変数が連続変数である任意のデータに対して無条件で適用できるわけではない。しかし、 $R1$ および $R2$ を一つの地域 $R12$ にまとめたとき、多くのマクロ経済変数については $R1$ と $R2$ の値を集計することで $R12$ の値を得ることができる。例えば、人口や金額などの量的なデータについては二つの地域の合計値が $R12$ の値となる。また、比率データについてはその分子と分母について二地域の合計値を求め、それらの比を計算することで $R12$ の値が得られる。このように集計できる数値を説明変数として設定することで、仮想の

^{*3} Zhang et al. (2018) では、目的変数である y_1, y_2 の値はワンホットエンコーディング（ダミー変数）としている。

地域 $R12$ における説明変数の値を求めることができる。一方、 $R12$ の投入係数についても $R1$ と $R2$ の投入係数を集計して計算することが可能である。浅利・土居 (2016)が述べるように、図2の競争移輸入型産業連関表における中間投入は対象地域内の産業間における投入額・国内他地域から対象地域への産業間の移入額・海外から対象地域への産業間の輸入額の和である。すなわち、ある地域における第 i 産業から第 j 産業への中間投入について、対象地域内での投入額を $\hat{m}_{ij}$ 、国内他地域から第 j 産業への移入額を $\dot{m}_{ij}$ 、海外から第 j 産業への輸入額を $\tilde{m}_{ij}$ とすると

$$m_{ij} = \hat{m}_{ij} + \dot{m}_{ij} + \tilde{m}_{ij}$$

となる。地域 $R1$ におけるこれら中間投入を $(m_{ij}^{R1}, \hat{m}_{ij}^{R1}, \dot{m}_{ij}^{R1}, \tilde{m}_{ij}^{R1})$ と表し、地域 $R2$ についても同様に $(m_{ij}^{R2}, \hat{m}_{ij}^{R2}, \dot{m}_{ij}^{R2}, \tilde{m}_{ij}^{R2})$ とする。これら二地域の中間投入を足し合わせると

$$m_{ij}^{R1} + m_{ij}^{R2} = \hat{m}_{ij}^{R1} + \dot{m}_{ij}^{R1} + \tilde{m}_{ij}^{R1} + \hat{m}_{ij}^{R2} + \dot{m}_{ij}^{R2} + \tilde{m}_{ij}^{R2}$$

となる。この中間投入の合計は、 $R12$ の中間投入 m_{ij}^{R12} に等しいことを示そう。地域 $R12$ における地域内での投入額 $\hat{m}_{ij}^{R12}$ は

$$\begin{aligned} \hat{m}_{ij}^{R12} &= \hat{m}_{ij}^{R1} + \hat{m}_{ij}^{R2} + \dot{m}_{ij}^{R1} \text{のうちのR2からの投入額} \\ &\quad + \dot{m}_{ij}^{R2} \text{のうちのR1からの投入額} \end{aligned}$$

となる。また、国内他地域から地域 $R12$ への投入額 $\dot{m}_{ij}^{R12}$ は

$$\begin{aligned} \dot{m}_{ij}^{R12} &= \dot{m}_{ij}^{R1} \text{のうちのR2を除く国内他地域からの投入額} \\ &\quad + \dot{m}_{ij}^{R2} \text{のうちのR1を除く国内他地域からの投入額} \end{aligned}$$

であり、また、海外から地域 $R12$ への投入額 $\tilde{m}_{ij}^{R12}$ は

$$\tilde{m}_{ij}^{R12} = \tilde{m}_{ij}^{R1} + \tilde{m}_{ij}^{R2}$$

となる。 $R12$ の中間投入額 m_{ij}^{R12} は $\hat{m}_{ij}^{R12}$ 、 $\dot{m}_{ij}^{R12}$ 、および $\tilde{m}_{ij}^{R12}$ の合計である

から、

$$\begin{aligned}
 m_{ij}^{R12} &= \hat{m}_{ij}^{R12} + \dot{m}_{ij}^{R12} + \tilde{m}_{ij}^{R12} \\
 &= \hat{m}_{ij}^{R1} + \hat{m}_{ij}^{R2} \\
 &\quad + \dot{m}_{ij}^{R1} \text{のうち } R2 \text{ からの投入額} \\
 &\quad + \dot{m}_{ij}^{R2} \text{のうち } R1 \text{ からの投入額} \\
 &\quad + \dot{m}_{ij}^{R1} \text{のうち } R2 \text{ を除く国内他地域からの投入額} \\
 &\quad + \dot{m}_{ij}^{R2} \text{のうち } R1 \text{ を除く国内他地域からの投入額} \\
 &\quad + \tilde{m}_{ij}^{R1} + \tilde{m}_{ij}^{R2} \\
 &= \hat{m}_{ij}^{R1} + \hat{m}_{ij}^{R2} + \dot{m}_{ij}^{R1} + \dot{m}_{ij}^{R2} + \tilde{m}_{ij}^{R1} + \tilde{m}_{ij}^{R2} \\
 &= m_{ij}^{R1} + m_{ij}^{R2}
 \end{aligned}$$

となる。したがって、任意の二地域について中間投入を合計することで、それら二地域を集計し新たな一つの地域とみなすときの中間投入が得られる。一方、集計した地域 $R12$ における第 j 産業の生産額 Y_j^{R12} は、 $R1$ と $R2$ の生産額の合計 $Y_j^{R1} + Y_j^{R2}$ に等しい*4。この結果、任意の二地域を集計した新たな地域 $R12$ の投入係数 a_{ij}^{R12} は、

$$a_{ij}^{R12} = (m_{ij}^{R1} + m_{ij}^{R2}) / (Y_j^{R1} + Y_j^{R2}) \quad (2)$$

となる。

*4 $R12$ の最終需要については、中間投入と同様に $R1$ と $R2$ の最終需要の和として計算できる。また、 $R12$ の輸出と輸入の差（純輸出）は、 $R1$ と $R2$ の純輸出の和となる。 $R12$ における移出と移入の差（純移入）については、追加の説明が必要であろう。二つの地域を一つの地域にまとめる際、 $R12$ における産業 i の移出 L_i^{R12} は

$$\begin{aligned}
 L_i^{R12} &= L_i^{R1} + L_i^{R2} - \sum_j \dot{m}_{ij}^{R2} \text{のうち } R1 \text{ からの投入額} \\
 &\quad - \sum_j \dot{m}_{ij}^{R1} \text{のうち } R2 \text{ からの投入額} \\
 &\quad - F_i^{R2} \text{のうち } R1 \text{ からの移入額} \\
 &\quad - F_i^{R1} \text{のうち } R2 \text{ からの移入額}
 \end{aligned}$$

となる。 L_i^{R1} は $R1$ における産業 i の移出、 L_i^{R2} は $R2$ における産業 i の移出、 F_i^{R1} と F_i^{R2} はそれぞれ $R1$ と $R2$ における産業 i の最終需要である。一方、 $R12$ における産業 i の移入

先述のmixupと異なり、上記のデータ拡張は三つ以上の地域についても成立する。したがって、二つ以上の地域についてそれらの説明変数および投入係数の値をそれぞれ集計することで、それら地域を一つの地域とみなした場合の説明変数と投入係数の値を求めることができる。

マクロ経済変数および投入係数に関する上記の性質を利用し、実在する地域を集計した仮想の地域を考えることでデータ拡張が実現できる。先ほどと同様に2015年の都道府県別のデータを用いるとしよう。このとき、それら都道府県の中から複数の県をランダムに抽出し上述の計算手順を使うことで、抽出した県を一つの地域とみなす場合の投入係数とその説明変数の値が生成できる。この生成を繰り返すことで、投入係数と説明変数についてのデータ拡張が可能になるといえる。

データ拡張における都道府県の抽出方法は、モデルの学習結果に影響することに注意すべきである。予測の観点から、データ拡張でデータを生成する際は多数の都道府県を抽出することは避けることが望ましい。式(2)より、抽出される都道府県が多いほど投入係数の値はすべての都道府県を集計した値に近づくことが分かる。説明変数に含まれる比率の値についても同様である。つまり、多くの都道府県を抽出し集計したデータは投入係数や一部の説明変数の分散が小さくなる。このように生成したデータは互いに非常に似た

N_i^{R12} は

$$\begin{aligned} N_i^{R12} = & N_i^{R1} + N_i^{R2} - \sum_j \dot{m}_{ij}^{R1} \text{のうち } R2 \text{ からの投入額} \\ & - \sum_j \dot{m}_{ij}^{R2} \text{のうち } R1 \text{ からの投入額} \\ & - F_i^{R1} \text{のうち } R2 \text{ からの移入額} \\ & - F_i^{R2} \text{のうち } R1 \text{ からの移入額} \end{aligned}$$

となる。 N_i^{R1} は $R1$ における産業 i の移入、 N_i^{R2} は $R2$ における産業 i の移入である。以上より、 $R12$ の産業 i の純移出は

$$L_i^{R12} - N_i^{R12} = L_i^{R1} + L_i^{R2} - N_i^{R1} - N_i^{R2}$$

となり、 $R1$ と $R2$ の純移出の和となる。生産額は、中間投入の行和 (= 中間需要計)・最終需要・純移出・純輸出の和に等しい。したがって、 $R12$ の生産額 Y_j^{R12} は $Y_j^{R1} + Y_j^{R2}$ と計算できる。

値をとるため、モデルの学習において新たな情報をほとんどもらさない。その結果、データ拡張により大きなデータを生成したとしても、多くの都道府県を集計したデータの割合が大きければ、データが小さい時と同様に過学習の問題をもたらしまう。

今回の深層学習に際しては、説明変数の選択についても注意が必要であると考え。都道府県のデータにより学習したDNNモデルを使って市町村の投入係数を予測する場合、市町村別に公表されないデータを説明変数として設定することはできない。例えば、都道府県ごとの全体的な所得は県民所得として公表されているが、市町村について同様のデータはほとんど公表されていない。このようなデータを説明変数に含めると、市町村を対象とする投入係数の推論はほぼ不可能となる。また、説明変数として数量を表すデータを含めることも避ける必要がある。例えば、人口を説明変数に含めたとしよう。上記のデータ拡張の過程で、複数の都道府県のデータが集計されることから、データ拡張により得た仮想的な地域の人口は都道府県の人口よりもさらに大きくなる。仮想的な地域の人口を説明変数に含めて学習したモデルは人口規模の大きい地域を表現するため、都道府県や市町村レベルの地域を対象とする予測には適切とはいえない。

今回のデータ拡張によって得られる新たな説明変数と目的変数の値の組は、都道府県の中から一部をランダムに抽出しそれらを一つの地域とみなすことで得たものである。その新たな地域はデータ拡張のためにまとめられた以上の意味を持たない。今回の深層学習の目的は、式(1)における投入係数の生成過程 f_{ij} を推定することである。分析に必要なデータは、説明変数の値とそれに対応する正しい投入係数の値の組であれば良く、それらの値の組が意味のある地域のものである必要はない。例えば、香川県・徳島県・愛媛県・高知県を集計した場合、その新たな地域は一般に四国地方という意味を持つ。一方、沖縄県・北海道・千葉県を集計して得た新たな地域は特段の意味を持たないであろう。しかし先に述べた通り、集計した地域を新たな一つの地域とみなすならば、いずれの説明変数・投入係数の組も学習に用いるデータと

して有効である。

4 モデルの学習と推論

これまでの考察に基づいて、実際に深層学習を行い投入係数の予測を試みる。はじめに、学習に利用するデータについて説明する。前節で考察した通り、都道府県のデータを用いて深層学習を行う。具体的には、2021年2月末時点で2015年産業連関表が公表されている34道府県のデータを用いる。投入係数は各都道府県の2015年産業連関表より計算する。また、今回の分析では産業を12部門に分類し各部門間の投入係数を計算する。したがって、投入係数行列の大きさは $12 \times 12 = 144$ となる。具体的な産業分類は、農林漁業、鉱業、製造業、建設、電力・ガス・水道、商業、金融・保険・不動産、運輸・郵便、情報通信、公務、サービス、分類不明の12部門とする。公表される産業連関表の産業分類は都道府県によって異なるため、各都道府県の2015年産業連関表を上記の12部門に再集計している。

これら34道府県に対して、各地域における人的資源の配分状況を示す産業別従業員割合および産業別事業所割合と、各地域における産業の生産性を示す事業所当たり売上額および事業所当たり付加価値額を計算し、それらを説明変数として用いる。産業別従業員割合および産業別事業所割合は、「平成26年経済センサス - 基礎調査」における産業（小分類）従業者数および事業所数の産業別割合として計算する。事業所当たり売上額および事業所当たり付加価値額は、「平成24年経済センサス - 活動調査」の産業（大分類）売上（収入）金額・付加価値額・事業所数より計算する^{*5}。これら説明変数の総数は1321となる。

前節のデータ拡張の手順に従い、34道府県からランダムに道府県を抽出し投入係数と説明変数を集計する。前節で述べたように、多くの道府県を抽出することは予測精度の点で望ましくない。今回のモデル学習では、34道府県

^{*5} これらのデータは、e-Stat の API 機能を使って収集した。

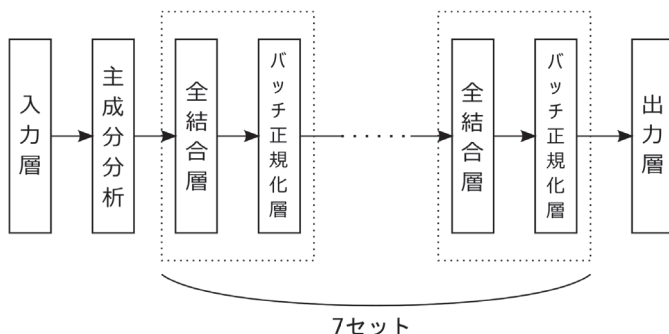


図 3: 学習に用いる DNN モデルの構造

から5以下の道府県を抽出する。実際には、5個の整数(1,2,3,4,5)について、34道府県から道府県を取り出す組み合わせ数 ${}_{34}C_1, {}_{34}C_2, {}_{34}C_3, {}_{34}C_4, {}_{34}C_5$ をウェイトして乱数をつつ生成し、その値の数だけ道府県をランダムに抽出する。抽出された道府県について投入係数と説明変数を計算し、それを一つの観測値とする。これを50000回繰り返してデータを生成する。このデータのうち、40000サイズを学習用に、10000サイズをテスト用にそれぞれ用いる。

次に、DNNモデルの形状と学習方法について説明する。学習に用いるモデルを図3に示している。入力となる説明変数を標準化した後で主成分得点に変換し新たな入力とする。このとき、寄与率の高い上位30位までの主成分得点を用いる。その後、512個のノードからなる全結合層とバッチ正規化層の組を7組連結する。最後のバッチ正規化層からの出力をシグモイド関数により結合し、最終的な出力を得る。モデル学習における誤差関数として平均二乗誤差を使い、最適化の手法としてバッチサイズを32とするミニバッチ勾配降下法を用いる。最適化の各ステップにおいて、学習用データの20%を検証用データに確保する。検証用データから計算した誤差関数の値が減少する間は最適解の探索を行い、誤差関数が増加傾向に転じたところで探索を打ち切

る。以上の設定に基づいてモデルを学習した後、テスト用データによるモデルの平均二乗誤差・平均絶対誤差を確認し、学習したモデルの妥当性を判断する*6。

最後に、実際のモデル学習と推論について説明する。12×12の大きさからなる投入係数行列の各要素を目的変数として深層学習を行う。ただし、34道府県のすべてについて投入係数の値が0となる要素については、それ以外の地域においてもその投入係数の値が0であるとみなし、モデルの学習を行わず予測値を0とする。したがって、実際に予測対象となる投入係数は130個であり、これらの投入係数ごとにDNNモデルを構築し学習する。学習したDNNモデルを用いて、学習用データに含まれない山口県および岐阜県郡上市の投入係数行列を推測し、モデルの予測精度について検証する*7。郡上市は、産業連関表の導出を独自に行っている数少ない市町村の一つである。今回のモデル学習では道府県のデータを用いているが、そのモデルがより小規模の自治体における投入係数の予測にも適用できるかを確認するため、郡上市を対象に投入係数を予測しその精度を確認する。

モデル学習および推論ではTensorflowを用いた。具体的には、Pythonのtensorflowパッケージを使い、モデル学習及び推論に関する計算を行った。また、Tensorflowでのモデル学習における計算を高速化するため、CUDA経由によるGPUサポートを利用している*8。

山口県の2015年産業連関表の投入係数行列について、表1にDNNモデルによる予測値を、表2に公表値を示している。また、郡上市の2015年産業連関表の投入係数行列について、表3にDNNモデルの予測値を、表4に公表値をそれぞれ示している。表5のDNNの列には、深層学習による予測値と公表されてい

*6 DNN モデルの学習に関する各種設定については、Géron (2019) を参考にしている。

*7 上述の通り、2021年2月時点で2015年産業連関表を公表している道府県のデータが学習用データのベースとなっているが、山口県の2015年産業連関表は2021年3月に公表されたため、山口県のデータは学習用データには含まれていない。

*8 なお、e-Stat APIによるデータ収集、収集したデータの結合、12産業分類への変換、データ拡張などの作業ではF#を利用した。また、説明変数のもととなるデータの選択は、収集したデータを目視で確認して行った。そのデータ選択用のGUIツールをC#により作成した。

表 1: 山口県 2015 年産業連関表投入係数行列 (深層学習による予測値)

	農林漁業	鉱業	製造業	建設	電力・ガス・水道	商業
農林漁業	0.0886	2.5968×10^{-5}	0.021	0.0014	0	0.0001
鉱業	1.5544×10^{-5}	0.0023	0.1162	0.0101	0.2332	2.5767×10^{-6}
製造業	0.2179	0.1261	0.4399	0.2687	0.0728	0.0341
建設	0.0023	0.0045	0.0022	0.0006	0.0122	0.0029
電力・ガス・水道	0.008	0.0439	0.0275	0.0042	0.1575	0.0262
商業	0.0675	0.0285	0.0392	0.0542	0.0184	0.011
金融・保険・不動産	0.0083	0.0477	0.0083	0.0166	0.0193	0.0396
運輸・郵便	0.0447	0.1776	0.0266	0.0455	0.0332	0.0403
情報通信	0.0045	0.0064	0.0057	0.0089	0.0100	0.0341
公務	0	0	0	0	0	0
サービス	0.0253	0.0634	0.032	0.1018	0.0573	0.0882
分類不明	0.0049	0.0117	0.0034	0.0119	0.0029	0.0075

表 1 の続き

	金融・保険・ 不動産	運輸・郵便	情報通信	公務	サービス	分類不明
農林漁業	2.0504×10^{-6}	0.0001	0	4.2002×10^{-5}	0.0068	0
鉱業	1.51×10^{-7}	3.3354×10^{-6}	2.7868×10^{-9}	1.0478×10^{-5}	1.2206×10^{-5}	0.0003
製造業	0.0105	0.1229	0.0336	0.0493	0.1348	0.0621
建設	0.0082	0.0048	0.0045	0.0083	0.0025	3.1454×10^{-5}
電力・ガス・ 水道	0.0045	0.0115	0.0108	0.0196	0.0265	0.0089
商業	0.0026	0.0244	0.0103	0.0102	0.0436	0.0115
金融・保険・ 不動産	0.0869	0.0443	0.0224	0.0196	0.0197	0.0306
運輸・郵便	0.0124	0.1391	0.0265	0.0347	0.0233	0.0872
情報通信	0.0184	0.0112	0.1874	0.0303	0.0224	0.0720
公務	0	0	0	0	0	0.2642
サービス	0.0501	0.1103	0.1683	0.1267	0.0909	0.0698
分類不明	0.0024	0.0102	0.004	0.0011	0.0052	0.0001

表 2: 山口県 2015 年産業連関表投入係数行列 (公表値)

	農林漁業	鉱業	製造業	建設	電力・ガス・水道	商業
農林漁業	0.1031	4.1918×10^{-5}	0.0124	0.0011	0	0.0002
鉱業	0.0001	0.0034	0.1098	0.007	0.2878	3.0884×10^{-6}
製造業	0.2070	0.0675	0.4367	0.2787	0.0877	0.0318
建設	0.0019	0.0021	0.0023	0.0006	0.0142	0.0029
電力・ガス・水道	0.0084	0.0356	0.0331	0.0034	0.0893	0.0284
商業	0.0593	0.0168	0.0321	0.0537	0.0189	0.0099
金融・保険・不動産	0.0089	0.033	0.0069	0.0164	0.0211	0.0343
運輸・郵便	0.0522	0.1932	0.0237	0.0436	0.0357	0.0454
情報通信	0.004	0.0065	0.005	0.0087	0.0121	0.0369
公務	0	0	0	0	0	0
サービス	0.0231	0.0368	0.0254	0.1013	0.0635	0.0894
分類不明	0.0047	0.012	0.0022	0.0137	0.0031	0.0067

表 2 の続き

	金融・保険・ 不動産	運輸・郵便	情報通信	公務	サービス	分類不明
農林漁業	2.1635×10^{-6}	1.5134×10^{-5}	0	3.2276×10^{-5}	0.0057	0
鉱業	0	6.0535×10^{-6}	0	7.4483×10^{-6}	1.5804×10^{-5}	0.0002
製造業	0.0103	0.1206	0.0374	0.0445	0.1307	0.0576
建設	0.0072	0.0051	0.0053	0.0091	0.0023	0
電力・ガス・水道	0.0042	0.0098	0.0105	0.0148	0.0204	0.0058
商業	0.0024	0.0243	0.0086	0.0096	0.0412	0.0101
金融・保険・不動産	0.0838	0.0458	0.0168	0.0211	0.018	0.0283
運輸・郵便	0.0123	0.178	0.0245	0.0332	0.0214	0.0946
情報通信	0.0198	0.0108	0.208	0.0296	0.0214	0.0755
公務	0	0	0	0	0	0.493
サービス	0.0522	0.104	0.1453	0.1252	0.0803	0.0633
分類不明	0.0024	0.0096	0.0032	0.0009	0.0045	0

表 3: 岐阜県郡上市 2015 年産業連関表投入係数行列 (深層学習による予測値)

	農林漁業	鉱業	製造業	建設	電力・ガス・水道	商業
農林漁業	0.1481	0.0001	0.0318	0.0016	0	0.0002
鉱業	2.5164×10^{-5}	0.0047	0.0133	0.0097	0.0452	1.8557×10^{-6}
製造業	0.2369	0.0387	0.4025	0.2565	0.0629	0.0358
建設	0.0024	0.0034	0.0026	0.0006	0.021	0.0037
電力・ガス・水道	0.0083	0.0299	0.0238	0.0029	0.1535	0.0265
商業	0.059	0.007	0.0543	0.0518	0.0074	0.0113
金融・保険・不動産	0.0037	0.0275	0.0085	0.0144	0.0293	0.0304
運輸・郵便	0.0704	0.2278	0.028	0.0559	0.0157	0.0878
情報通信	0.0034	0.0067	0.008	0.0084	0.0243	0.0353
公務	0	0	0	0	0	0
サービス	0.0158	0.0248	0.0379	0.1144	0.2622	0.0861
分類不明	0.0033	0.0088	0.0037	0.0135	0.0059	0.0076

表 3 の続き

	金融・保険・ 不動産	運輸・郵便	情報通信	公務	サービス	分類不明
農林漁業	2.6483×10^{-6}	0.0001	0	3.0486×10^{-5}	0.0078	0
鉱業	2.955×10^{-7}	1.7429×10^{-6}	1.2857×10^{-8}	5.4371×10^{-6}	1.3758×10^{-5}	0.0005
製造業	0.0142	0.1691	0.041	0.0485	0.1411	0.1194
建設	0.0086	0.0075	0.0056	0.0126	0.0028	7.0136×10^{-6}
電力・ガス・水道	0.0028	0.007	0.012	0.0145	0.0235	0.0075
商業	0.0027	0.0476	0.0078	0.0109	0.0473	0.0115
金融・保険・不動産	0.065	0.0307	0.0127	0.0179	0.017	0.0372
運輸・郵便	0.0176	0.0788	0.0282	0.0431	0.0334	0.1445
情報通信	0.0229	0.0082	0.2064	0.0323	0.0183	0.0722
公務	0	0	0	0	0	0.2556
サービス	0.0542	0.1724	0.1229	0.1233	0.0741	0.0735
分類不明	0.0021	0.0064	0.0036	0.0007	0.0057	2.276×10^{-5}

表 4: 岐阜県郡上市 2015 年産業連関表投入係数行列 (公表値)

	農林漁業	鉱業	製造業	建設	電力・ガス・水道	商業
農林漁業	0.1336	0	0.033	0.0018	0	0.0002
鉱業	0.0001	0	0.0014	0.0121	0	3.7924×10^{-6}
製造業	0.1971	0.1446	0.3697	0.2366	0.0796	0.0352
建設	0.0018	0.0012	0.0018	0.0005	0.0301	0.0032
電力・ガス・水道	0.014	0.2058	0.0241	0.0031	0.0923	0.0349
商業	0.0601	0.0256	0.0542	0.0463	0.017	0.0105
金融・保険・不動産	0.0099	0.023	0.01	0.0182	0.0249	0.0394
運輸・郵便	0.0767	0.1624	0.0455	0.0598	0.0246	0.0699
情報通信	0.0028	0.0013	0.0063	0.0092	0.0396	0.0405
公務	0	0	0	0	0	0
サービス	0.0188	0.0234	0.0341	0.1347	0.1465	0.0992
分類不明	0.0055	0.0042	0.0046	0.0087	0.0091	0.0074

表 4 の続き

	金融・保険・ 不動産	運輸・郵便	情報通信	公務	サービス	分類不明
農林漁業	2.0304×10^{-6}	1.5675×10^{-5}	0	1.2372×10^{-5}	0.0078	0
鉱業	5.076×10^{-7}	1.3631×10^{-6}	0	2.911×10^{-6}	1.6245×10^{-5}	0.0002
製造業	0.0091	0.1427	0.0205	0.0348	0.1104	0.0577
建設	0.0079	0.0054	0.0043	0.0109	0.0026	0
電力・ガス・水道	0.0026	0.0365	0.0121	0.0146	0.0416	0.0053
商業	0.0021	0.0404	0.005	0.0093	0.0385	0.0101
金融・保険・不動産	0.0771	0.0347	0.0214	0.0192	0.0281	0.0344
運輸・郵便	0.0126	0.0687	0.0272	0.0392	0.0269	0.1032
情報通信	0.0164	0.0028	0.2273	0.0282	0.0181	0.0755
公務	0	0	0	0	0	0.2466
サービス	0.0406	0.1886	0.1507	0.1243	0.0807	0.0634
分類不明	0.0017	0.0017	0.0032	0.0008	0.0056	0

表 5: 予測精度の各種指標

	山口県		郡上市	
	DNN	RAS	DNN	RAS
STPE	0.1497	0.143	0.262	0.2753
MAD	0.0059	0.0056	0.0099	0.0104
U ₂	0.2657	0.1617	0.34	0.519
RMSE	0.0219	0.0133	0.0241	0.0368
MAPE	0.1945	0.2696	0.2246	0.0818

る値との乖離を測るための予測誤差の指標をまとめている。これらは、産業
 連関表間の推定誤差・予測誤差の測定に際して Hosoe (2013) が用いている指
 標である。STPEは標準誤差率、MADは平均絶対差、U₂はTheilのU₂、RMSE
 は平均平方誤差、MAPEは平均絶対誤差率で、それぞれ以下のように計算さ
 れる。

$$\begin{aligned}
 \text{STPE} &= \sum_{i,j} |\tilde{a}_{ij} - a_{ij}| / \sum_{i,j} a_{ij} \\
 \text{MAD} &= \sum_{i,j} |\tilde{a}_{ij} - a_{ij}| / N_a \\
 \text{U}_2 &= \sqrt{\sum_{i,j} (\tilde{a}_{ij} - a_{ij})^2} / \sqrt{\sum_{i,j} a_{ij}^2} \\
 \text{RMSE} &= \sqrt{\left\{ \sum_{i,j} (\tilde{a}_{ij} - a_{ij})^2 \right\} / N_a} \\
 \text{MAPE} &= (1/N_a) \sum_{i,j} |(\tilde{a}_{ij} - a_{ij}) / a_{ij}|
 \end{aligned}$$

ただし、 $\tilde{a}_{ij}$ は投入係数の予測値、 a_{ij} は公表されている投入係数の値、 N_a は投
 入係数の個数である。予測値と公表値の差が大きいほどこれら指標の値は大
 きくなる。深層学習による投入係数の予測精度をRAS法と比較するため、表5
 のRASの列に、RAS法による予測値について同様の指標を示している。RAS

法で用いる投入係数行列の初期値は、山口県の場合は全国の2015年産業連関表の投入係数行列を、郡上市の場合は岐阜県の2015年産業連関表の投入係数行列を、それぞれ用いている。また、各産業の生産額・中間投入・中間需要として実際の公表値を使用している。したがって、表5のRAS法による予測では各産業の生産額・中間投入・中間需要の予測誤差は0であり、最良の予測値を用いているといえる。実際のRAS法による予測では、これらの予測誤差は0とならないことから、表5に示す指標は変動しうる。

深層学習による予測精度について考察しよう。表1と表2を概観すると、山口県の各投入係数について、深層学習の予測値は公表値に総じて近い値をとっていることが分かる。ただし、公務から分類不明への投入係数のみ公表値と予測値との間に大きな乖離がある。この投入係数について、34道府県の公表値が0.2程度である一方、山口県の公表値は0.493であり山口県の値が際立って高い。その原因は不明であるが、DNNモデルは34道府県の投入係数に基づいて学習が行われたため、この投入係数の値を正しく予測できなかったといえる。表5の予測精度の指標をみると、STPEとMAEはDNNモデルとRAS法とでほぼ同じ大きさと、 U_2 とRMSEはRAS法が小さく、MAPEはDNNモデルが小さい。したがって、これらの指標からはDNNモデルとRAS法で予測精度の優劣を判定することが難しい。しかし、先述の通り、RAS法による予測では各産業の生産額・中間投入・中間需要の予測誤差を0としていることから、実際のRAS法では表5の値は変動する。そのため、実際のRAS法と比較して、深層学習による方法はより安定した精度で投入係数を予測できているといえる。

郡上市についても、表3と表4より、深層学習による予測値は公表値に近い値をとっているようである。また、表5の予測精度の指標について、STPEとMAEはDNNモデルとRAS法でほぼ同じ大きさとなり、 U_2 とRMSEはDNNモデルが小さく、MAPEはRAS法が小さい。そのため、郡上市についても、DNNモデルとRAS法の予測精度は同程度である。

以上の予測結果より、少なくとも山口県と郡上市について、深層学習によ

る予測値は総じて公表値に近く、各産業の生産額・中間投入・中間需要の予測誤差を0とした場合のRAS法の予測値とほぼ同程度の予測精度を持つことが明らかとなった。

5 考察

本研究では、複数の地域を集計して仮想的な一地域とみなすことで投入係数と説明変数のデータ拡張を行い、そのデータによる深層学習を通じて投入係数を予測する方法を考察した。実際に34道府県のデータに基づいて深層学習を行い、山口県と郡上市の投入係数を予測した。その結果、山口県と郡上市について、深層学習による予測の精度は、各産業の生産額・中間投入・中間需要に公表値を用いたRAS法の予測精度とほぼ同等であることがわかった。実際のRAS法では、各産業の生産額・中間投入・中間需要の値に予測誤差が含まれることから、深層学習に基づく方法はRAS法よりも安定した精度での予測が可能であると考ええる。また、道府県レベルのデータで学習したDNNモデルであっても、郡上市の投入係数を比較的高い精度で予測できていることは興味深い。

今回学習したモデルは、都道府県の投入係数の推計過程を説明するものである。前節の深層学習では34道府県のデータを利用したが、これら道府県の中間投入および投入係数は厳密には推計値である。したがって、学習したDNNモデルは実際には投入係数の推計過程を説明するものであり、一次統計に基づく産業連関表の作成過程を直接説明するものではない。一次統計から産業連関表を作成し計算した投入係数を真の投入係数とするならば、深層学習で予測される投入係数は、真の投入係数の予測値ではなく、真の投入係数の推測値の予測値である。

RAS法と比較して、深層学習による投入係数の予測には多くの利点がある。第一に、学習したモデルは他の地域の投入係数の予測に再利用できる。今回の分析では実際に、34道府県のデータにより学習したDNNモデルをそ

のまま山口県と郡上市の予測に用いている。当然、他の県や市町村についても説明変数の値が用意できれば、同じDNNモデルを使ってそれら地域の投入係数が予測できる。第二に、RAS法と異なり、投入係数を予測する際に初期時点の投入係数行列を用意する必要がなく、各産業の生産額・中間投入・中間需要についての予測も不要である。RAS法の計算手順自体は非常に単純であるが、その実際の計算では初期時点の投入係数行列を用意し、各産業の生産額・中間投入・中間需要の予測を行わなければならない。特に、各産業の生産額・中間投入・中間需要の予測誤差は投入係数の予測精度に影響することから、RAS法ではこれらの値についても高い精度で予測することが要求される。一方、深層学習による予測では、各産業の生産額・中間投入・中間需要を予測する必要はない。したがって、深層学習ではこれらの値の予測に伴う手間を省くことができる。第三に、モデル改良により予測精度を向上する余地が多分に残されている。今回のDNNモデルで設定した説明変数は、各産業の生産性を表す比較的単純な変数であった。実際には、投入係数の生成過程における説明変数として、他にも様々な変数が挙げられる。新たな説明変数を導入することにより、さらに予測精度の高いモデルを構築することができるだろう。あるいは、モデルの改良や学習方法の改善を通じても予測精度を向上できると考える。ただし後述するように、説明変数として利用できる変数には一定の制限が課される点には注意が必要である。

一方、深層学習の利用には一定程度の計算資源が要求される。上記のモデル学習では、12産業部門における投入係数行列の各要素について予測を行った。これら投入係数のうち34道府県のすべてでその値が0でないものを予測の対象としたため、実際には130個の投入係数が予測対象となる。つまり、今回の投入係数行列の予測では、130個のDNNモデルの学習が必要であった。今回の分析では比較的安価なデスクトップPCを用いたが、学習の開始から完了までに約2日を要した。仮に、統合大分類である37産業部門について今回と同じコンピュータを用いてDNNモデルで投入係数を予測する場合、産業部門数が約3倍となることから、計算には $2日 \times 3^2 = 18日$ 程度が必要と予想

する。中分類や小分類の投入係数行列を予測するならば、計算にかかる時間はさらに増大するだろう。しかし見方を変えれば、十分な計算資源を用意することで、細分化された産業部門の場合でもDNNモデルによる投入係数の予測は実現可能であるといえる。例えば、37産業部門における予測の場合、同じ性能のコンピュータを9台用意できれば2日程度でモデルの学習が可能となる。あるいは、高い性能を持つコンピュータを利用することで、さらに短い時間でモデルの学習を完了できるだろう。したがって、十分な計算資源を保有しているならば、深層学習の導入に伴う計算時間の増大はそれほど大きな問題ではないといえる。

また、予測する対象に応じて説明変数の設定が制約されることも注意しなければならない。現状では、説明変数に量的データを含めることは望ましくない。また、市町村の投入係数を予測するならば、都道府県のみ公表されるデータを説明変数に設定することはできない。

上述の通り、深層学習により、山口県と郡上市に対して比較的高い精度で投入係数を予測できることを示した。しかし、深層学習の予測精度を厳密にとらえるためには、さらに多くの地域について投入係数を予測する必要がある。学習用データのもととなる34道府県のうち、いくつかの道府県のデータを予測精度を確認するために確保するという対策が考えられる。しかし、産業連関表が公開されている地域が少ない現状では、産業連関表が利用できる地域のデータは可能な限り学習用データに含めることが望ましい。一方、上記の分析で観測対象から除外した13の県については、これから2015年産業連関表が公開されるだろう。そこで今後の研究では、今回の手法を用いてこれら13県の投入係数を予測し、深層学習による予測精度をさらに精密にとらえることを目指す。

本研究は、産業連関表のうち投入係数部分の推計のみに着目している。任意の地域についての産業連関表を推計するためには、投入係数以外の項目についても推計する必要がある。本研究の手法により投入係数行列を推計した場合、各産業の生産額・中間投入・中間需要・最終需要のいずれか一つが推計

できれば、それ以外の項目を計算することができる。ただし、付加価値および最終需要の詳細な項目に関しては、別途推計する必要がある。投入係数行列の深層学習による推計を通じた産業連関表全体の推計手法については、今後の研究で構築を目指す。

本研究で提示したデータ拡張の方法は、他のマクロ経済指標についても適用できる可能性がある。その適用方法が確立できるならば、様々なマクロ経済指標について深層学習により精緻な予測が可能になるだろう。

■謝辞 山口大学経済学部に加藤真也先生には、分析に利用するデータや予測精度の計測方法などに関して多くの助言をいただいた。ここに感謝の意を表する。

参考文献

- Bishop, C. M. (1995) “Training with Noise is Equivalent to Tikhonov Regularization,” *Neural Computation*, Vol. 7, No. 1, pp. 108-116.
- Géron, Aurélien (2019) *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*: O'Reilly, 2nd edition.
- Hosoe, N. (2013) “Estimation Errors in Input-Output Tables and Prediction Errors in Computable General Equilibrium Analysis,” GRIPS Discussion Paper 13-19, National Graduate Institute for Policy Studies.
- McCulloch, W. S. and W. H. Pitts (1943) “A Logical Calculus of the Ideas Immanent in Nervous Activity,” *Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115-133.
- McDougall, R. A. (1999) “Entropy Theory and RAS are Friends,” GTAP Working Papers 6.
- Papadas, C. T. and W. G. Hutchinson (2002) “Neural Network Forecasts of Input-Output Technology,” *Applied Economics*, Vol. 34, pp. 1607-1615.

- Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986) “Learning Internal Representations by Error Propagation,” *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, pp. 318-362.
- Wu, S., H. R. Zhang, G. Valiant, and C. Ré (2020) “On the Generalization Effects of Linear Transformations in Data Augmentation,” *arXiv:2005.00695*.
- Zhang, H., M. Cisse, Y. N. Dauphin, and D. Lopez-Paz (2018) “Mixup: Beyond Empirical Risk Minimization,” *arXiv:1710.09412*.
- 朝日幸代 (2004) 「平成7年名古屋市産業連関表の作成の試み」, 『産業連関』, 第12巻, 第1号, 16-24頁.
- 浅利一郎・土居英二 (2016) 『地域間産業連関分析の理論と実際』, 日本評論社.
- 岩崎俊夫 (1980) 「産業連関分析と経済予測：RAS方式による投入係数修正の妥当性について」, 『北海道大學経済学研究』, 第30巻, 第1号, 121-142頁.
- 長谷川雅志 (1991) 「RAS法における誤差の推計結果の統計的分布への影響について」, 『地域学研究』, 第22巻, 第1号, 185-204頁.
- 日吉拓也・河上哲・土井正幸 (2004) 「ノンサーベイ・アプローチによるつくば市産業連関表の作成と応用」, 『産業連関』, 第12巻, 第1号, 3-15頁.