

質的選択行動のニューラルネットワーク モデルによる分析

Analysis of Qualitative Choice Behavior by Neural Network Model

釜 国 男

Kunio KAMA

- | | |
|-------------------|---------|
| 1. はじめに | 4. 予測結果 |
| 2. 階層型ニューラルネットワーク | 5. むすび |
| 3. 計測結果 | |

1. はじめに

経済的な選択はしばしば二者択一の形をとる。たとえば家計は家屋を保有するかしないか、新車を購入するかしないか、就業するか失業するか等の選択を行う。また経済以外の分野でも、選挙における投票行動や学生の進路の決定、住所の選択などその例は無数にある。近年、個々の家計や企業に関するサーベイ・データが比較的容易に利用できるようになり、それらのデータを分析するためにいくつかの質的選択モデルが開発されている。そのうち最も広く使われているのは、線型確率モデル、プロビットモデル、ロジットモデル、およびトービットモデルである。このうち線型確率モデルは、通常の回帰方程式で被説明変数がただ2つの値だけをとるようにしたもので、計算は簡単であるが実際に適用するには多くの問題がある。なかでも大きな欠点は、推定したモデルからの予測値が負となったり1より大きくなる場合があるという問題であ

る。この難点を回避して確率が(0,1)の範囲内になるように工夫したのが、プロビットモデルとロジットモデルである。いま、経済主体がある行動を選択する確率を P_i で表すと、これらのモデルでは

$$P_i = F(X_i \beta)$$

で与えられる。ここで X_i は説明変数で、 β はその係数ベクトルである。 F は確率分布関数を表し、プロビットモデルでは標準正規分布関数を使用し、ロジットモデルでは

$$F(z) = 1 / (1 + \exp(-z))$$

という累積ロジスティック関数を用いる。これらのモデルはパラメータに関して非線形であり、最尤法を用いて推定される。一方プロビット分析を拡張したトービットモデルでは、事象の生起だけでなくその程度も扱うことが可能である。たとえば新車を購入するかしないかの選択に加えて、購入した場合の支出額も知りたいようなときにこのモデルは有用である。

このように質的選択についてはすでにいくつかの統計的なモデルが考案されているが、

本稿では同じ問題をまったく別の観点から分析する方法を提示し、既存の方法と比較してその有効性を示す。すなわち脳による情報処理のモデルであるニューラルネットワークを構築し、これにバックプロパゲーション法を適用して変数間の関係を学習させ、それをもとに未学習のデータを予測するという新しい手法を提案する。そして実際のデータに適用して、ニューラルネットワークの優れた学習能力と予測力を実証し、提案した方法が質的データの分析にきわめて有効であることを示す。

2. 階層型ニューラルネットワーク¹⁾

ニューラルネットワークは脳の神経組織を模した人工的なシステムであり、システムを構成するのはニューロンと呼ばれる多入力、

1) ニューラルネットワークの原理と応用については、中野・飯沢〔5〕、麻生〔1〕、甘利〔3〕など優れた解説書が少なくない。また安居院・長橋・高橋〔2〕にはBP学習法のコンピュータプログラムが掲載されており、パソコンで楽しみながら学ぶことができる。

1 出力の素子である(図1)。各入力シナプスに相当する重みがつけられてニューロンに入力され、ニューロンはそれらの総和としきい値から

$$f(x) = 1 / (1 + \exp(-x))$$

というシグモイド関数をもとに出力値を計算する²⁾。重みは興奮性の結合のときは正の値となり、抑制性の結合のときは負の値をとる。またシナプス結合の可塑性に対応して、重みは学習によって変化させることができる。

ニューラルネットはこうした多数のニューロンが結合してできた回路網であり、一つのニューロンの情報はシナプスを通じて別のニューロンに伝わる仕組みになっている。そしてニューロンどうしのダイナミックな相互作用によって並列の情報処理を行う。ニューラルネットワークは大別して、それぞれの

2) これは先に示した累積ロジスティック関数そのものである。関数は0～1の範囲内の値をとるので、出力値を事象の確率として解釈するのに都合がよい。この他にも、出力値が0か1の値だけをとるマカロックとピッツのモデルや、確率モデル、カオスモデルなどがあるが、質的な選択行動のモデルには適していない。

図1 ニューロンモデル

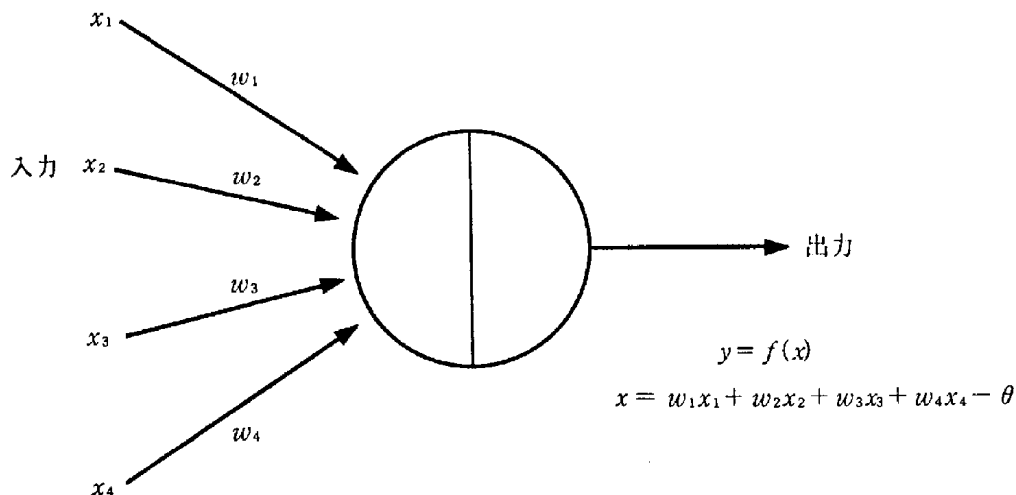
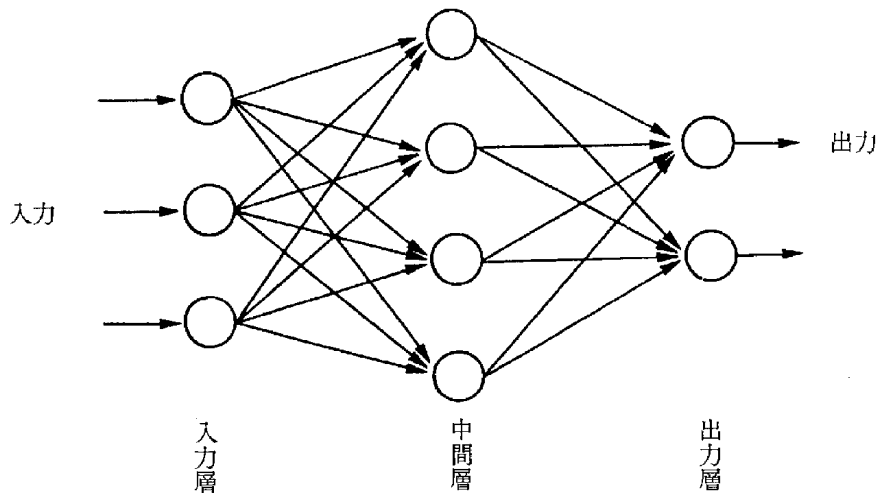


図2 3層ニューラルネットワーク



ニューロンがシナプスを介して他のすべてのニューロンと結合している相互結合型のネットワークと、ニューロンが層状にグループ化され信号がグループ間を特定の方向にのみ伝わる階層型のネットワークとに分けられる。相互結合型ではホップフィールド・ネットワークが有名であり、階層型ではローゼンブラットの提唱したパーセプトロンや多層ネットワークが代表的である。このうち本研究で用いるのは3層ネットワークであり、図2のような構造を持っている。ニューロンは入力層、中間層、出力層の3層に配置され、入力層へ入力された外部からの情報は中間層で変換されてから出力層に入り、ここで最終的な処理を受けて出力される。同じ層のニューロン間では情報のやり取りは行わない。一般にニューロン間の結合係数は可変であり、結合係数を変化させることによって同じネットワークを様々な情報処理に適したものに変えることができる。この過程をネットワークの学習と呼んでいる。問題はいかにして学習させるかであるが、パーセプトロンのような階層型ネットワークには、ラメルハート [6] らの考案したバックプロパゲーション法 (誤

差逆伝播法; BP法) が広く使われている。これはネットワークの出力信号と、外部から与えられる正答の教師信号を比較して、デルタルールにもとづいて結合係数の大きさを徐々に変えてゆく一種の確率的な探索法である。いま

x_k : 入力

y_j : 中間層のニューロンの出力

z_i : 出力層のニューロンの出力

t_i : 教師信号

とすると、 z_i と t_i の差が小さいほど望ましい。そこで2乗誤差の総和

$$E = \sum (z_i - t_i)^2 \quad (1)$$

が最小となるように結合係数を調整する。そのために学習パターンを1個入力するたびに、中間-出力層間の結合係数 w_{ij} と入力-中間層間の結合係数 w_{jk} 、およびしきい値の γ_j 、 θ_k を以下のように修正する。

$$\Delta w_{ij} = \alpha \cdot \delta_i \cdot y_j$$

$$\Delta \gamma_j = \beta \cdot \delta_j, \quad \delta_i = (t_i - z_i) \cdot z_i \cdot (1 - z_i)$$

$$\Delta w_{jk} = \alpha \cdot \sigma_j \cdot x_k \quad (2)$$

$$\Delta \theta_k = \beta \cdot \sigma_k, \quad \sigma_j = \sum \delta_i \cdot w_{ji} \cdot y_j \cdot (1 - y_j)$$

ここで、 α と β は学習係数である。学習係数が小さすぎると極小解に到達するのに時間が

かかり、逆に大きすぎると極小解を飛び越し学習が振動して収束しないことがあるので、適当な値に設定する必要がある。この学習法がバックプロパゲーション（逆伝播）と呼ばれる由縁は、通常の答えを出すための信号は入力層から出力層へと流れていくのに対して、誤差を修正するための信号は逆に出力層から入力層へと流れていくからである。なお、(1)式で表される誤差曲面には複数の極小値が存在して初期値によってはローカルミニマムに落ち込むことがある。ここではこの問題を回避するために、初期値をランダムにとって2万回の学習を行い、これを10回繰り返して誤差が最小となるケースをグローバルミニマムないしそれに近い解として採用することにした。また、計算にあたって入力データの一部をあらかじめ適当に変換してモデルの精度を高めるように工夫した。

一般に階層型ニューラルネットワークは1個以上の中間層を持っているが、特殊なケースとして中間層のない入力層と出力層だけからなるモデルを考えると、これは先に紹介したロジットモデルそのものとなる。すなわちしきい値と結合係数をそれぞれ定数項および回帰係数と解釈すると、1個の出力を持つ3層のニューラルネットワークは伝統的なロジットモデルを一般化したものと見ることができる。

3. 計測結果

本節では、3層ネットワークを実際のデータに当てはめた結果を示す。使用したデータは、ピンダイック・ルビンフェルド〔7〕の“Econometric Models and Economic Fore-

casts”の第10章に載っている学校教育にたいする支出計画への賛否に関するクロスセクション・データである。データは以下の通りである。

YESVM = 賛否を表すダミー変数で賛成なら1, 反対ならば0

PUB12 = 1, 2名の子供が公立学校に在学なら1, そうでないとき0

PUB34 = 3, 4名の子供が公立学校に在学なら1, そうでないとき0

PUB5 = 5名以上の子供が公立学校に在学なら1, そうでないとき0

PRIV = 子供が私立学校に在学なら1, そうでないとき0

YEARS = 学区に住んでいる年数

SCHOOL = 教師として雇用されているなら1, そうでないとき0

LINC = 年収の対数（自然対数）

PTCON = 年間の学校税支払い額の対数（自然対数）

データは全部で95個あるが、そのうち1～75番目を学習用に用い、残りの76～95番目のデータを予測のために使用する。ニューラルネットとの比較のために、先ずはじめに伝統的な3つのアプローチを適用しよう。

(1) 線型確率モデル

YESVMを被説明変数、PUB12以下の変数を説明変数とする回帰式を推定すると、結果は次のようになる。（括弧内はt値である）

$$\begin{aligned} \text{YESVM} = & -1.0793 + 0.1805\text{PUB12} \\ & (-0.57) \quad (1.05) \\ & + 0.3667\text{PUB34} - 0.2802\text{PUB5} + 0.02475\text{PRIV} \\ & (1.96) \quad (-0.76) \quad (0.12) \\ & - 0.002919\text{YEARS} + 0.2643\text{SCHOOL} \\ & (-0.49) \quad (1.41) \end{aligned}$$

$$+0.3926LINC-0.3543PTCON$$

$$(2.45) \quad (-1.59)$$

$$\widehat{R}^2=0.215, SE=0.467$$

決定係数は0.215と低く、年収以外の変数はすべて統計的に有意ではない。推定式から計算した理論値を確率として解釈すると、いくつかのケースで計画案に賛成の確率がマイナスとなったり、1より大きくなる。また0.5を境に二つのグループに分類すると、22のケースで誤って分類してしまう。こうした点からみて、線型確率モデルを適用するのは難しい。

(2) プロビットモデル

次のようなモデルを推計した。(YESVM=1は賛成, YESVM=0は反対を示す)

$$\text{Prob}(\text{YESVM}=1) = G(\beta_1 + \beta_2\text{PUB12} + \beta_3\text{PUB34} + \beta_4\text{PUB5} + \beta_5\text{PRIV} + \beta_6\text{YEARS} + \beta_7\text{SCHOOL} + \beta_8\text{LINC} + \beta_9\text{PTCON}) \quad (3)$$

最尤法で推定すると、結果は次のようになる。

$$\beta_1 = -6.8381 \quad \beta_4 = -7.0056 \quad \beta_7 = 1.4998$$

$$(-1.12) \quad (0.00) \quad (1.48)$$

$$\beta_2 = 0.5444 \quad \beta_5 = 0.1732 \quad \beta_8 = 1.6342$$

$$(1.08) \quad (0.28) \quad (2.68)$$

$$\beta_3 = 1.1684 \quad \beta_6 = -0.003038 \quad \beta_9 = -1.4294$$

$$(2.08) \quad (-0.18) \quad (-1.82)$$

5%水準で統計的に有意となるのは、PUB34とLINCの2つの変数だけである。モデルの精度をチェックするために、YESVMとその理論値 $\widehat{\text{YESVM}}$ から決定係数に似た統計量

$$\widehat{R}^2 = 1 - \text{ESS}/\text{TSS}$$

$$\text{ESS} = \sum (\text{YESVM}_i - \widehat{\text{YESVM}}_i)^2$$

$$\text{TSS} = \sum (\text{YESVM}_i - \overline{\text{YESVM}})^2 \quad (4)$$

を計算すると、 $\widehat{R}^2=0.755$ となる。またパターン分類のモデルとして解釈すると、53個のケースでパターンを正しく分類し、残りの22

ケースでは誤って分類する。したがって統計的なフィットという点では、このモデルは一応合格点をつけることができる。

(3) ロジットモデル

次にロジットモデルを当てはめると、係数推定値は次のようになる。

$$\beta_1 = -11.4624 \quad \beta_4 = -22.1524 \quad \beta_7 = 2.5037$$

$$(-1.09) \quad (0.00) \quad (1.42)$$

$$\beta_2 = 0.8512 \quad \beta_5 = 0.2322 \quad \beta_8 = 2.6772$$

$$(1.04) \quad (0.23) \quad (2.61)$$

$$\beta_3 = 1.8589 \quad \beta_6 = -0.005851 \quad \beta_9 = -2.2957$$

$$(2.02) \quad (-0.20) \quad (-1.73)$$

$\widehat{R}^2=0.762$ で、パターン分類の問題としてみると54のケースで正答となり、プロビットモデルより説明力はやや高くなる。

(4) ニューラルネットワーク

つづいて同じデータセットを用いてニューラルネットワークを同定した。ネットワークの入力、中間、出力層はそれぞれ8, 12, 1個のニューロンから構成され、バックプロパゲーション法の調整係数は $\alpha = \beta = 0.8$ とした。また入力データのうち、学区に住んでいる年数、年収、学校税支払い額は次式により0~1の範囲内にスケールする(生のデータのままだでは学習が困難である)。

$$\tilde{x} = 1 / (1 + \exp(-(x - \bar{x})))$$

ここで x は入力データ、 \bar{x} はその平均値である。表1は2万回の学習で得られた結合係数の値を示している。ネットワークの構造はN(8,12,1)であるので、しきい値まで含めると係数は全部で121個ある。

つぎに、このネットワークを用いて事後的予測を行うと $\widehat{R}^2=0.833$ で、従来法に較べて誤差は小さくなる。また出力が0.5以上であるときは賛成、未満のときは反対に分類する

表 1 結合係数

	入 力 層								しきい値	
	1	2	3	4	5	6	7	8		
中 間 層	1	-2.456	-0.225	-1.267	0.920	-1.871	0.074	-3.004	-1.008	-1.209
	2	-1.687	-1.169	-0.113	-1.115	0.549	-0.462	-0.715	-0.555	-1.992
	3	8.100	-0.999	0.654	-0.015	-10.537	-1.372	-12.496	-11.073	6.785
	4	-9.135	-7.261	0.475	-7.000	1.878	0.435	31.009	-20.165	0.149
	5	-0.872	-1.480	-0.558	-0.864	0.066	0.611	-2.268	-5.035	-2.108
	6	-5.161	-7.613	2.744	4.545	6.307	-0.359	-37.808	15.194	11.245
	7	-0.565	6.316	-1.768	7.359	-14.632	-2.792	-0.339	-16.690	2.432
	8	3.921	-0.680	1.814	1.884	4.579	-1.037	-19.999	-12.376	7.057
	9	-1.860	-5.178	2.429	0.458	-1.778	-2.406	-20.110	3.550	5.731
	10	-14.402	-9.918	0.511	-3.317	42.831	1.255	14.522	-24.398	-19.586
	11	-4.144	3.103	-0.830	3.009	-3.090	6.462	-1.865	-13.028	4.533
	12	13.747	5.479	-1.512	1.053	-8.466	4.556	-71.317	37.839	10.302

〔中間層－出力層〕

$$\begin{array}{llll}
 w_1=0.965 & w_4=22.624 & w_7=16.000 & w_{10}=23.085 \\
 w_2=-0.067 & w_5=0.689 & w_8=-9.970 & w_{11}=8.584 \\
 w_3=-7.444 & w_6=-22.216 & w_9=-8.035 & w_{12}=23.957 \\
 \gamma=-16.145 & & &
 \end{array}$$

と、誤って分類するのは4つのケースに限られ、正答率は $71/75=0.95$ となる。したがって同定したネットワークは確率モデルとしてよりも、むしろ二者択一の分類モデルとして使用するのが望ましい。いずれにしてもニューラルネットワークの方が学習用データの説明力は高いが、これはニューロンの動作関数としてシグモイド関数を使用している関係で、入力と出力の間に強い非線形性が存在するためである。この点を示したのが図3である。年収、学校税支払い額と賛否投票の理論値との関係を示すこの図で、他の説明変数は $PUB12=0$, $PUB34=1$, $PUB5=0$, $PRIV=1$, $YEARS=8.98$ (サンプル平均値), $SCHOOL=0$ に固定してある。年収を一定とすると、計画に賛成する確率は学校税支払い額に応じて複雑に変化し、また学校税が同じでも年収によって確率は大きく異なる。図4はロジスティックモデルから導出した入出力

関係を示している。この場合には、年収を一定とすると賛成する確率は学校税が増えると低くなり、また学校税が同じなら年収の高い世帯ほど計画に賛成する確率が高くなる。2つの図を比較すると、中間層を加えることでニューラルネットの近似精度が飛躍的に高まることがわかる。すなわち中間層のない2層のネットワークであるロジスティックモデルの入出力曲面はごく単調に変化するのに対して、中間層を加えて3層にすると入出力曲面は非常に複雑な形状を持つようになる。同じ非線形のモデルでも、データに含まれる複雑な非線形関係を正確に表現するには3層のネットワークにする必要がある。

さらに中間層のニューロンの個数を変えて、同じデータにたいして再度学習を行わせた。スペースの都合で同定されたネットワークの誤差 (R^2) だけを記すと、中間層のニューロンが、6個:0.831, 8個:0.828, 10個:

0.835, 14個：0.833, 16個：0.833, 18個：
0.835, 20個：0.834となる。N(8,12,1)のネット
ワークでは $\hat{R}^2=0.833$ であるので、中間層
のニューロンを増減しても誤差はほとんど変
わらないことがわかる。一般に中間層の

ニューロンを増やすことによって、より複雑
な関数を近似できるようになるが、増やしす
ぎると過学習 (overfitting) がおきて未学習
データに対する予測精度が悪くなる。した
がって中間層には12個程度のニューロンを配

図3 年収、学校税支払い額と賛成の確率 (ニューラルネットワーク)

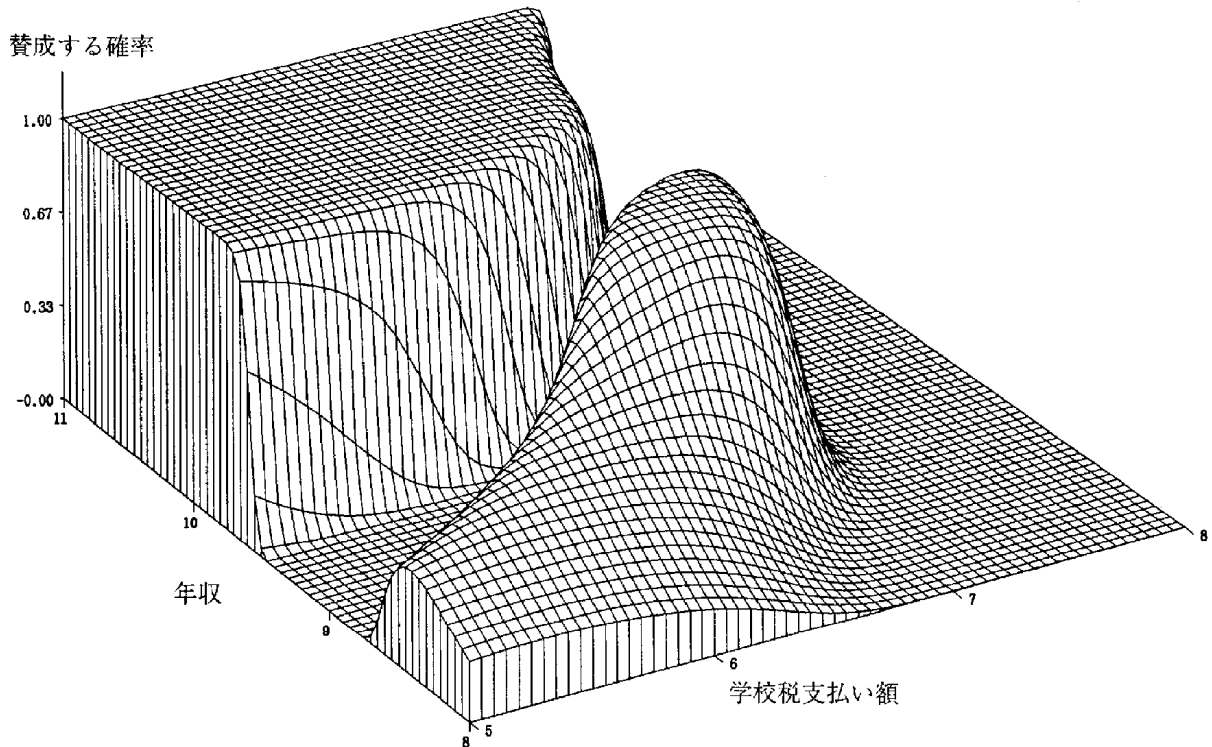
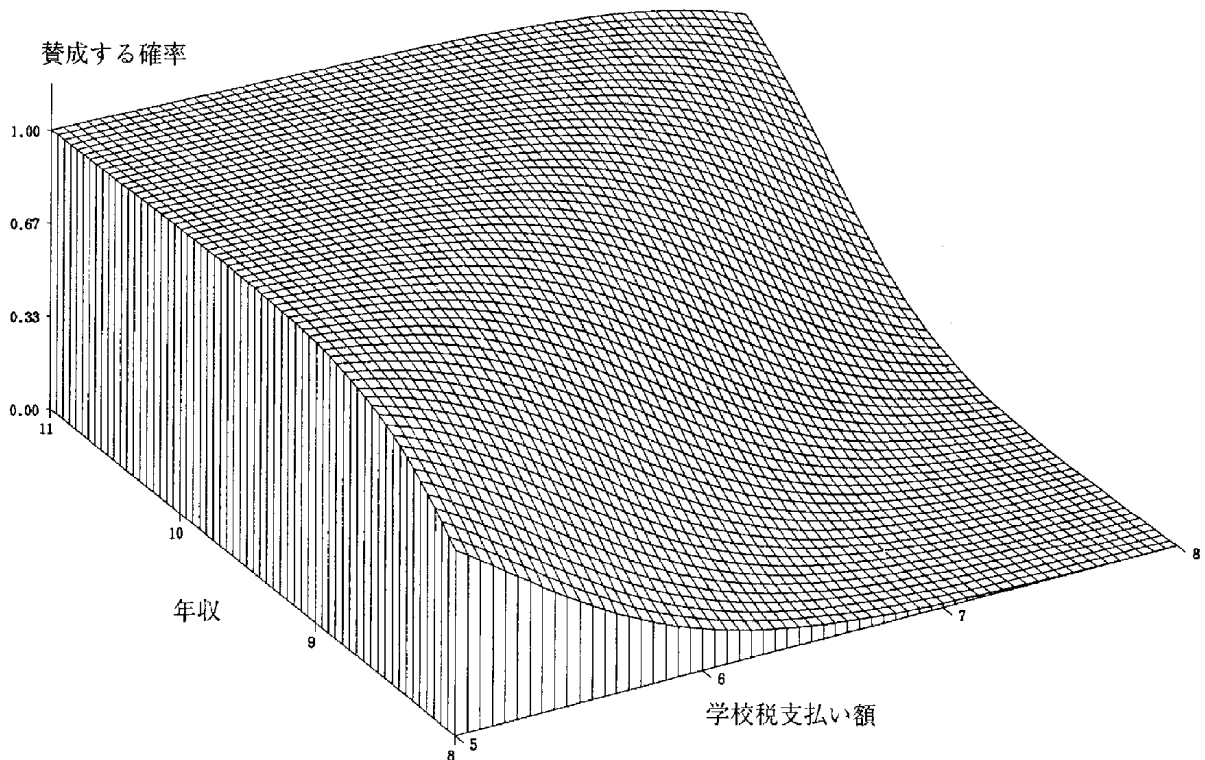


図4 年収、学校税支払い額と賛成の確率 (ロジットモデル)



すれば十分であろう。

4. 予測結果

最後に推定したモデルを用いて未学習データの予測を行った。予測は3つの伝統的なモデルとニューラルネットワークを用いて行い、表2はその結果をまとめたものである。この表で予測の数値は各世帯が支出計画に賛成する確率を表している。4つの方法を比較すると、従来法による予測値は0と1の中間の値をとる傾向があるのに対して、ニューラ

ルネットによる予測は1に近い確率に集中している。表の数値から予測誤差の自乗和を計算すると、OLS：5.438、Probit：5.636、Logit：5.592、NN：4.581となり、ニューラルネットによる誤差が最も小さい。また0.5未満を0、以上を1とみなすと、OLSとProbitでは11ケース、Logitでは12ケースで正答となるのに対して、NNでは15ケースで正答が得られる。したがって賛成・反対への分類問題としてみた場合にも、やはりニューラルネットの方が従来の予測法より優れていると言える。

表2 未学習データに対する予測結果

No.	実績値	OLS	Prob	Logit	NN
1	1	0.891	0.915	0.903	1.000
2	1	0.811	0.876	0.870	1.000
3	0	0.838	0.906	0.894	1.000
4	1	0.797	0.935	0.926	1.000
5	1	0.792	0.829	0.826	1.000
6	1	0.451	0.371	0.376	0.998
7	1	0.991	0.962	0.948	1.000
8	0	0.432	0.341	0.343	1.000
9	1	0.688	0.748	0.757	0.999
10	0	0.309	0.253	0.244	1.000
11	1	0.456	0.406	0.397	0.239
12	1	0.043	0.000	0.000	1.000
13	1	0.358	0.301	0.308	0.998
14	1	0.716	0.734	0.736	1.000
15	1	0.457	0.455	0.466	0.999
16	1	0.527	0.499	0.507	0.998
17	1	0.981	0.983	0.972	1.000
18	1	0.064	0.000	0.000	0.962
19	1	0.463	0.717	0.725	1.000
20	0	0.731	0.760	0.765	1.000

5. む す び

個々の家計や企業のマイクロデータを対象とする実証分析では、これまで主としてプロビットモデルその他の統計的な手法が用いられてきたが、本稿ではそれらに代わるものとしてニューラルネットワークに基づく新しい方法を提案した。この方法によると、モデルの関数型をあらかじめ特定する必要がなく、データに合わせて結合係数を調整するので非線形の複雑な関係でも自由に表現することが可能である。このために、3, 4節の結果から明らかなように、従来法に比べて学習用データに対するフィットは良好で、未学習データでも7割以上の正答率となる。したがって、ここで検討したデータに関する限り、提案した方法は従来の統計的な手法より優れていると言える。

質的データとしては、回答が3つ以上の選択肢を含む場合がある（たとえば、賛成、反対、および棄権の3つのうちからどれか1つを選択するような場合）。ニューラルネットワークでこうしたデータを分析するには、出力層のニューロンを複数個用意し、結合係数に適切な制約条件を付けるだけで十分である。しかるに既存の多肢選択モデルを推定するには複雑な非線形推定が必要であり、あまり実用的ではない。

さらに、工夫次第ではニューラルネットワークでも説明変数を有意なものだけに絞り込むことが可能である。そのひとつの方法として、市川〔4〕の再構築学習法がある。通常の学習過程のなかに忘却過程を取り入れて学習と消去を繰り返すこの学習法は、もともとネット

ワークの構造を簡素化するために考案されたものであるが、説明変数の選択にも利用できるであろう。

提案した方法はこのように多くの利点を持っているが、まだいくつか解決しなければならない問題が残されている。

- (1) 学習時間：一般にバックプロパゲーション法では学習が収束するまでに多数回の繰り返し計算が必要である。このためシステムの構造変化に合わせてネットワークを頻繁に調整したい場合には、計算アルゴリズムを改良して学習時間を短縮する必要がある。もっとも普通のコンピュータの上でニューラルネットのプログラムを走らせる代わりに、ニューロ専用のマシン（ニューロコンピュータ）を作ればこの問題は一挙に解決するであろう。ハード化の研究はすでに始まっており、その実用化が待たれる。
- (2) ローカルミニマムの問題：結合係数の初期値によってはローカルミニマムに陥り大域的な最小値に達しない場合がある。これを回避する方法として、学習過程に確率的なゆらぎを導入することが考えられる。
- (3) 学習パラメータの問題：バックプロパゲーションを実行するには、学習係数（およびシグモイド関数の傾斜）を適切に設定しなければならないが、いまのところ一般的な決め方は見つかっていない。問題ごとに試行錯誤で決めているのが実情であるが、何らかの目安でもあれば時間と労力を要するこの作業を大幅に節約できる。

これらは解決不可能な問題ではないので、

理論と応用面での研究が進めばいずれは解決されるものと思われる。

参 考 文 献

- [1] 麻生英樹『ニューラルネットワーク情報処理』産業図書, 1988年
- [2] 安居院猛・長橋宏・高橋裕樹『ニューロプログラム』昭晃堂, 1993年
- [3] 甘利俊一『神経回路網の数理』産業図書, 1978年
- [4] 市川紘『階層型ニューラルネットワーク』共立出版, 1993年
- [5] 中野馨・飯沼一元編『ニューロコンピュータ』技術評論社, 1989年
- [6] D. E. Rumelhart, et al, *Parallel Distributed Processing*, MIT Press, 1986 (甘利俊一監訳『PDPモデル』産業図書, 1989年)
- [7] R. S. Pindyck and D. L. Rubinfeld, *Econometric Models and Economic Forecasts*, McGraw-Hill, 1976 (金子敬生監訳『計量経済学(上・下)』マグローヒル好学社, 1981年)