

確率ニューラルネットを用いた 意思決定支援システム

奥原 浩之, 三上 大輔, 川畑 興求
広島県立大学 経営学部 経営情報学科

1. はじめに

一般に, 被説明変数を複数の説明要因によって説明するための手法としては, 重回帰分析を用いることが考えられるが, 線形モデルを仮定するためモデル化において大きな誤差が残る可能性がある。また, 重回帰分析では得られた結果を用いて被説明変数から各説明要因を説明するような逆問題を解くことはできない。これらの問題に対処するために, ここで提案するシステムは未知の非線形モデルを学習によって推定することができるニューラルネットワークを組み込んでいる。

近年, ニューラルネットワークはパターン認識や信号処理等の分野に適用され有効性が示されており, 種々のモデルが提案されている。提案するシステムはニューラルネットワークを組み込むことで非線形モデルに対応できるためモデル化において誤差を小さくできる。また, 逆問題を解く能力を備えたニューラルネットを採用することにより, 被説明変数の値といくつかの説明要因を与えることで, 着目する操作可能な説明要因をどのように操作すれば望ましい被説明変数の値が期待できるのかを示すことができるシステムとなっている。

2. データ例と従来法による解析結果

ここでは, 本研究の導入としてデータ例について説明し, 従来法である重回帰分析による分析結果について述べる。まず, データは一般的に得られる表 1 のようなものを考えた。

Table 1 実質民間最終消費支出に関するデータの例

年度-期	CP	CPI	GDP.N	INTPR	WAGLN
1970-1	25,109.5	31.7	16,173.9	8.2	13.7
1970-2	25,983.2	32.2	16,832.7	8.5	17.4
⋮	⋮	⋮	⋮	⋮	⋮
1998-4	74,738.1	103.1	134,015.7	2.2	123.8

ここで, 被説明変数は実質民間最終消費支出 (CP)[単位: 10 億円] であり, 操作不可能な説明要因として消費者物価指数全国総平均 (CPI)[単位: 1995 年度を 100%], 名目国内総生産 (GDP.N)[単位: 10 億円] ならびに全産業の名目賃金指数 (WAGLN)[単位: 1995 年度を 100%] があり, 長期プライムレート (INTPR)[単位: 年利%] が操作可能な説明要因である。データは期間が 1970 年から 1998 年の四半期ごとの日本経済のマクロデータの一部であり, 116 組からなる。

このようなデータが与えられた時に被説明変数を複数の説明要因によって説明する方法として, まず重回帰分析を用いることが考えられる。予測に用いるデータを線形モデル

$$Y = A\theta + e$$

で表す。ここで, $Y \in \mathbb{R}^n$ は確率ベクトル, $A \in \mathbb{R}^n \times \mathbb{R}^{p+1}$ は計画行列, $\theta = [\theta_0, \theta_1, \dots, \theta_p]^T \in \mathbb{R}^{p+1}$ は未知母数, $e \in \mathbb{R}^n$ は残差ベクトルである。ここで, 残差の二乗和である二乗誤差を $E = e^T e$ で表す。また, 行列 A の i , ($i = 1, 2, \dots, n$) 行 1 列要素は 1 である。つまり, θ_0 は切片に相当する。そこで, 表 1 のデータについては $n = 116$, $p = 4$ となる。データに対して重回帰分析した結果を表 2 に示す。

Table 2 未知母数の推定値

	θ_0	θ_1	θ_2	θ_3	θ_4	E
推定値	27296.7573	-82.3710	0.3977	-400.0452	24.7586	129338608.1000

ここで、推定値は小数点以下5桁目を四捨五入している。

この結果から、ある特定の説明要因が被説明変数にどのように影響しているのかといった傾向がある程度はつかむことはできる。しかしながら、線形モデルを当てはめたことにより二乗誤差がかなり残っているうえ、得られた結果を用いて被説明変数から各説明要因を説明するといった逆問題を解くことはできない。

そこで、非線形モデルを用いた推定を行うことでさらに二乗誤差を減少させ、逆問題を解く能力を備えさせることで、被説明変数の値といくつかの説明要因から着目する説明要因をどのように操作すれば望ましい被説明変数の値が得られるのかを示すことができるシステムを考える。

3. 提案するシステムの概要

本研究では、提案するシステムにニューラルネットワーク (Neural Networks: 以下, NN)[1] の一つである確率ニューラルネットワーク (Probabilistic Neural Network: 以下, PNN)[2] を用いることとする。PNNは未知の確率密度関数 $p(\mathbf{z})$ を複数の動径基底関数の足し合わせで推定するニューラルネットワークである。動径基底関数としては正規確率密度関数を用いられる。ここでは、 K 個の入力ニューロンと1個の出力ニューロンからなるPNNについて説明する。 $(N+M)$ 次元の第 s 入力ベクトル $\mathbf{z}_s \in R^{(N+M)}$, ($s=1, 2, \dots, S$) は全ての入力ニューロンに入力される。ここで、 S はサンプルされたデータの数である。第 k 入力ニューロン ($k=1, 2, \dots, K$) はパラメータ ϕ_k と荷重 w_k をもつ。パラメータ ϕ_k は平均ベクトルと共分散行列の集合 $\{\mathbf{m}_k, \Sigma_k\}$ であるものとする。ここで、 $\mathbf{m}_k = [m_k^1, m_k^2, \dots, m_k^{(N+M)}]^T$ であり、 Σ_k はその逆行列 Σ_k^{-1} の第 ij 要素に σ_k^{ij} をもつ $(N+M) \times (N+M)$ の行列である。また、 Σ_k は対角行列で正定値対称行列である。

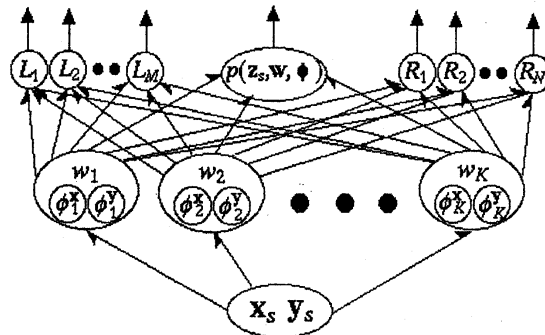


図1 PNNを組み込んだ提案システムの構造

いま、入力ベクトル $\mathbf{z}_s$ が $\mathbf{x}_s = [x_s^1, x_s^2, \dots, x_s^N]^T \in R^N$ と $\mathbf{y}_s = [y_s^1, y_s^2, \dots, y_s^M]^T \in R^M$ から $\mathbf{z}_s^T \equiv [\mathbf{x}_s^T, \mathbf{y}_s^T]$ で構成されているとする。また、入力ニューロンの動径基底関数においてベクトル $\mathbf{x}_s$ と $\mathbf{y}_s$ の間に

$$N_{(N+M)}(\mathbf{z}_s, \phi_k) = N_N(\mathbf{x}_s, \phi_k^x) \times N_M(\mathbf{y}_s, \phi_k^y)$$

の関係を仮定する。ここで、

$$N_{(N+M)}(\mathbf{z}_s, \phi_k) = \frac{1}{(2\pi)^{(N+M)/2} \|\Sigma_k\|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{z}_s - \mathbf{m}_k)^T \Sigma_k^{-1} (\mathbf{z}_s - \mathbf{m}_k)\right\}$$

であり、添字の T はベクトルの転置を示している。つまり、各入力ニューロンにはそれぞれ入力ベクトルのうち $\mathbf{x}_s$ を受け持つ細胞と $\mathbf{y}_s$ を受け持つ細胞から構成されることとなる。その結果、 $R_i = E_R[x_s^i | \mathbf{y}_s : \theta]$, ($i=1, 2, \dots, N$) を出力する N 個の右側出力ニューロン群と $L_i = E_L[y_s^i | \mathbf{x}_s : \theta]$, ($i=1, 2, \dots, M$) を出力する M 個の左側出力ニューロン群を考えることができる (図1 参照)。ここで、 $\mathbf{w}$ で集合 $\{w_1, w_2, \dots, w_K\}$, ϕ で集合 $\{\phi_1, \phi_2, \dots, \phi_K\}$, パラメータ θ で集合 $\{\mathbf{w}, \phi\}$ を表し、

$$E_R[\mathbf{x}_s | \mathbf{y}_s : \theta] = \int_{R^N} \mathbf{x} p(\mathbf{x} | \mathbf{y} : \theta) d\mathbf{x} = \sum_{k=1}^K \alpha_k(\mathbf{y}_s) \mathbf{m}_k^x$$

である。ただし、

$$\alpha_k(\mathbf{y}_s) = \frac{w_k N_M(\mathbf{y}_s, \phi_k^y)}{\sum_{k=1}^K w_k N_M(\mathbf{y}_s, \phi_k^y)}$$

である。左側出力ニューロン群の出力も同様に得ることができる [3]。

第 k 入力ニューロンは入力ベクトル $\mathbf{z}_s$ に対して

$$\xi(\mathbf{z}_s, w_k, \phi_k) = w_k N_{(N+M)}(\mathbf{z}_s, \phi_k)$$

を出力する。出力値 $\xi(\mathbf{z}_s, w_k, \phi_k)$ は中央出力ニューロンへ伝達され、中央出力ニューロンでこれらは足し合わされ

$$p(\mathbf{z}_s, \mathbf{w}, \phi) = \frac{1}{\sum_{k=1}^K w_k} \sum_{k=1}^K \xi(\mathbf{z}_s, \phi_k)$$

が出力される。このとき、パラメータ ϕ の各推定値は対数尤度関数

$$L(\mathbf{z}_s, \mathbf{w}, \phi) = \log p(\mathbf{z}_s, \mathbf{w}, \phi)$$

を最大化することにより得られる。これらの推定値は Expectation Maximization(以下, EM) アルゴリズム [4] を用いて

$$\begin{aligned} w_k^{(t+1)} &= \frac{1}{S} \sum_{s=1}^S h_k^{(t)}(\mathbf{z}_s) \\ m_k^{(t+1)} &= \frac{\sum_{s=1}^S \mathbf{z}_s h_k^{(t)}(\mathbf{z}_s)}{\sum_{s=1}^S h_k^{(t)}(\mathbf{z}_s)} \\ \Sigma_k^{(t+1)} &= \frac{\sum_{s=1}^S (\mathbf{z}_s - \mathbf{m}_k^{(t)})(\mathbf{z}_s - \mathbf{m}_k^{(t)})^T h_k^{(t)}(\mathbf{z}_s)}{\sum_{s=1}^S h_k^{(t)}(\mathbf{z}_s)} \end{aligned}$$

の反復計算により求めることができる。ここで、

$$h_k^{(t)}(\mathbf{z}_s) = \frac{w_k^{(t)} N_{(N+M)}(\mathbf{z}_s, \phi^{(t)})}{\sum_{k=1}^K w_k^{(t)} N_{(N+M)}(\mathbf{z}_s, \phi^{(t)})}$$

である。

対数尤度関数 $L(\mathbf{z}_s, \theta)$ を最大化するパラメータ θ を求めるために最急降下法を適用することも考えられるが [3], 学習係数の設定によって学習回数が増加することが予想される。そのため、提案するシステムでは EM アルゴリズムから導出された学習法を適用し、入力ニューロンと中央出力ニューロンが学習のために用いられる。そして、入力ニューロンで獲得されたパラメータ θ を利用して右側ならびに左側出力ニューロン群の出力を得る。このように、システムは入力ニューロンを通じて学習する部分と出力を得る部分が統合されたものとなっている。

4. 数値実験の結果と考察

4.1 提案システムによる学習結果

本研究において提案された PNN を用いたシステムを採取されたデータに適用する場合、入力ベクトル $\mathbf{z}_s$ は説明要因 4 つ (CPI, GDP, N, INTPR, WAGI, N) と被説明変数 1 つ (CP) から構成されることとなる。そのため、 $S = 116$, $\mathbf{x}_s$ の要素数は $N = 4$ となる。また、 $\mathbf{y}_s$ の要素数は $M = 1$ であり、システムに用意した入力ニューロンの数は $K = 300$ である。

図 2 (a) ~ (c) は表 1 で与えられたデータに対して、PNN が EM アルゴリズムを用いた学習により CP に影響を与えていると考えられる説明要因 ($\equiv x_s^i, i = 1, 2, \dots, 4$) と CP ($\equiv y_s$) の同時確率密度関数を獲得していく様子を示している。横軸は説明要因のうち CPI ($\equiv x_s^1$) を表し、縦軸は CP の値を表す。以後、図中の CP の単位は 10 兆円とする。× は CPI に対して観測された CP の値であり、● は学習過程における左側出力ニューロンの出力 L_1 である。図 2 (c) は学習終了時に PNN が獲得している説明要因 (CPI, GDP, N, INTPR, WAGI, N) と CP の同時確率密度関数を反映しており、図 2 (d) の重回帰分析の結果と比較して大幅に二乗誤差 E が減少していることがわかる。線形モデルを当てはめた重回帰分析と比較して、非線形モデルを仮定する PNN を用いたシステムによるモデル化は大幅に二乗誤差を減少させることができる。

4.2 提案システムによる要因分析

つぎに、提案したシステムが学習後の結果を用いて、特定の説明要因が被説明変数にどのように影響しているのかを説明できることを示す。ここで、図 3 においてシステムが獲得した確率密

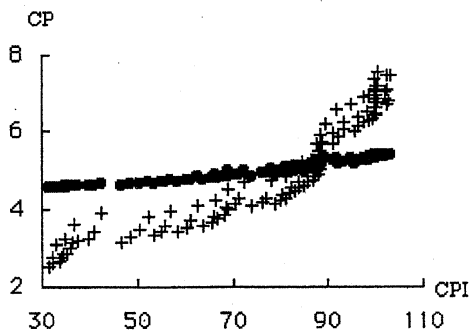


図 2 (a) PNN の初期状態
($E = 15011048209.34$)

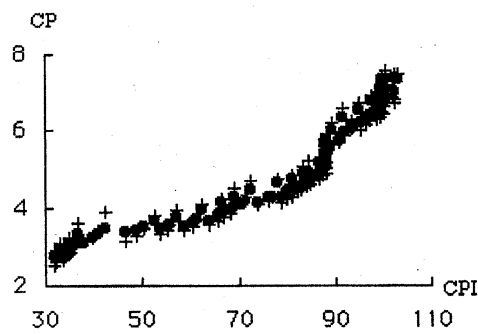


図 2 (b) PNN の過渡状態
($E = 212765941.68$)

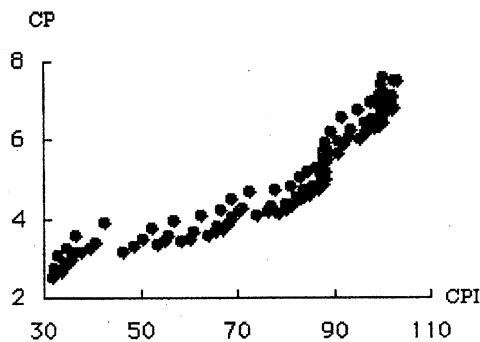


図 2 (c) PNN の定常状態
($E = 835664.26$)

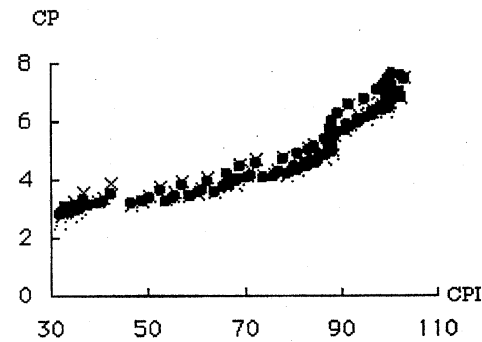


図 2 (d) 重回帰分析の結果
($E = 129338608.10$)

度関数 $p(\mathbf{z}, \theta)$ を用いるために、パラメータ ϕ のうち CPI と CP に関する部分のパラメータ ϕ_1 と ϕ_5 を入れ替えて、CPI を y_s へ入力し、CP を x_s^1 へ入力する。

図 3 (a) では横軸は説明要因のうち CPI ($\equiv y_s$) を表し、縦軸は CP の値を表す。やはり、 $\times$ は CPI に対して観測された CP の値であり、実線は CPI の最小値から最大値までの間を等間隔に 580 点に分割したものを y_s へ入力したときに得られた出力 R_1 の値を結んだものである。出力 R_1 は左側出力ニューロン群の出力 R_i ($i = 1, 2, \dots, 4$) のうち CP に関する出力となっている。同様に WAGLN の場合に対して得られた結果を図 3 (b) に示す。

図 3 (a) の結果から、CPI がおよそ 80% までとそれ以降で CP に対する影響の度合いが異なっていることがわかる。CPI が 80% より小さいときには CPI が 20% 増加する毎に CP が 5 兆円程度増加しているのに対して、80% をこえると CP は 30 兆円程度増加していることが示される。このことから CP は CPI が 80 ~ 90% の時期において急激に伸びていることがわかる。また図 3 (b) の結果から、WAGLN が 60% までと 100% 以降では出力の変動が小さいのに対して、60 ~ 100% の区間では大きな変動が見受けられる。これは同じような WAGLN の値に対して異なる CP の値がその区間において観測されていたからである。これらの考察は特定の説明要因のみを抽出することではじめて得られるものである。その他の説明要因 (GDP.N, INTPR) が CP に及ぼす影響についても線形モデルを当てはめた重回帰分析と比較して、非線形モデルを仮定する PNN を用いたシステムによるモデル化により詳しく分析することができる。

4.3 提案システムの意思決定支援への適用

次に、提案したシステムが逆問題を解く能力を備えたニューラルネットを採用しているため、被説明変数の値といくつかの説明要因を与えることで、着目する説明要因をどのように操作すれば望ましい被説明変数の値が得られるのかを各期ごとに提示できることを示す。ところで、表 2 の重回帰分析の結果からわかるように INTPR は CP に対して負の影響を与えているといえる。しかしながら、このような結果が得られたからといって、CP を高くする (あるいは低くする) ためにすべての状況に対して同じ方策 (INTPR を低くする、あるいは高くする) を施すことがよいとはいえない。各期の CPI, GDP.N, WAGLN の状況を考慮した上で判断するのが適当である。

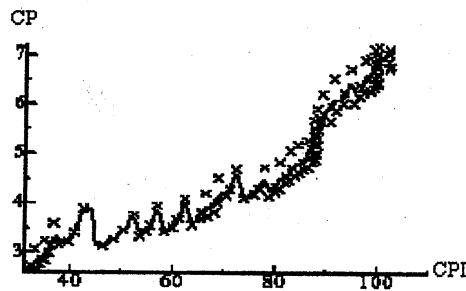


図3 (a) CPIとCPの関係

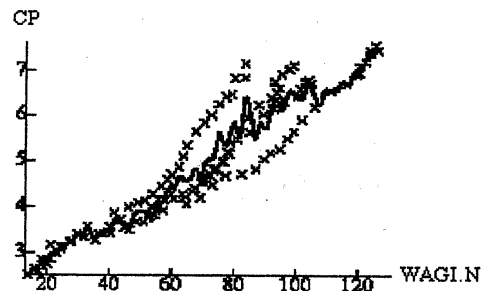


図3 (b) WAGI.NとCPの関係

そこで、操作可能な説明要因INTPRを y_s 、CPを x_s^3 としてPNNに学習させ、学習後にCP($\equiv x_s^3$)を変化させて左側出力ニューロン L_1 の出力を観測する。このとき、その他の説明要因の値は x_s^j ($j \neq 3$)は着目する1番目のデータ($s=1$)の値で固定されている。図4(a)は1977年の第3-四半期に着目したときに得られた結果である。図4(a)の横軸はCPの値を表し、縦軸は操作可能な説明要因($\equiv y_s$)の値を表す。また、実線はCPの最小値から最大値までの間を等間隔に580点に分割したものを x_{26}^3 へ入力したときに得られた出力 L_1 を結んだものである。同様に1998年の第4-四半期に対して得られた結果を図4(b)に示す。

グラフが途中で途切れているのは、その期の残りの説明要因(CPI,GDP.N,WAGI.N)の同じような値に対してCPがほとんど観測されていないことを意味しており、システムの出力 L_1 は0となっている。つまり、過去において観測の例がない場合は判断を留保することがこのシステムでは可能である。また、点線の先の●はその期に観測されたCPとCPIの値に対応している。図4(a)では(39819.3,9.2)、図4(b)では(74738.1,2.2)にプロットされている。

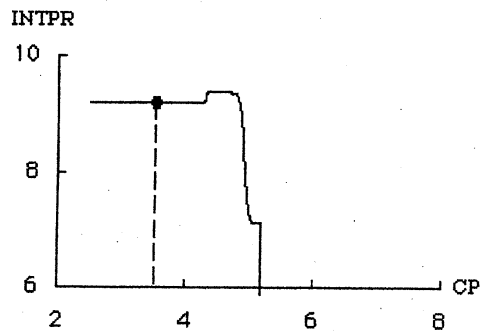


図4 (a) 1977年の第3-四半期の場合

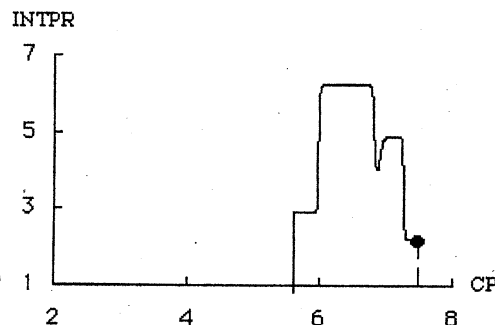


図4 (b) 1998年の第4-四半期の場合

図4(a)の結果から、1977年の第3-四半期に対しては操作可能である説明要因INTPRを9.2%に設定しているため、CPとしては45兆円程度までの値のどれかが得られていた可能性があることがわかる。また、この期の状況でCPを50000円にしたければINTPRを8%周辺で設定すればよいという方策が導かれる。さらに、このときのCPI,GDP.N,WAGI.Nと同じような値では、採取されたデータにおいてCPが52兆円を超えた例が存在しなかったか、存在しても非常に数少なかったことがわかる。つぎに図4(b)の結果は、1998年の第4-四半期に設定されたINTPRがもし3~4%の間であれば、確率的にCPが60兆円あるいは77兆円周辺のどちらかの値をとっていた可能性が高いことを示している。このことは、この期のCPI,GDP.N,WAGI.Nと同じような値に対して、採取されたデータにおいて二通りのCPの値が得られていたことを示している。このような場合は、INTPRの設定のみではCPが一意に定まらないケースであり、より詳しい分析が必要であることがわかる。これらの結果から、提案したシステムでは望ましいCPの値を得るために操作可能な説明要因INTPRに対する適切な方策を各期ごとに示唆できることが示された。

4.4 提案システムによる予測への適用結果

最後に、本研究で提案したシステムが予測に用いることが可能であることを、1996年第4-四半

期から1998年第4四半期までの8期分をテストデータとして予測することにより示す。予測の手法はまず、1970年第1四半期から1996年第3四半期のデータを学習させる。そのときに、 $n+1$ 期と n 期のCPの値の差をとり、適当な定数を足すことによりCPの値をCPの伸びを反映する正の値に変換した。データを変換したのはCPの値が1996年第3四半期までは説明要因に対して増加あるいは減少する傾向を示していることから、提案システムで予測を行う場合、1996年第4四半期以降のデータを未知のデータであるとして0を出力し続けるからである。ここでは、テスト期間（8期分）のCPを一括して予測する場合と、テスト期間の着目する期のデータが観測されたという条件の下で次の期のCPを予測する場合を考えて比較した。

図5(a)に一括してテスト期間全てのCPを予測した結果を示し、図5(b)に次の期ごとのCPを予測した結果を示す。実線は観測されているCPの値である。点線がシステムが予測したCPの値である。図5(a),(b)の比較から伸び率の誤差が積み重なっていくため不利であると考えられる一括予測と逐次予測の結果が大きくは異ならないことがわかる。この結果は、提案したシステムが採取されたデータの母集団の確率密度関数を推定しているためであると考えられる。

しかしながら、ともにテスト期間の2期目である1997年第1四半期のCPの予測値は大きく観測値からはずれている。その原因は提案したシステムが期待値を出力するためである。つまり、予測のために用いられた1996年第4四半期の説明要因(CPI,GDP.N,INTPR,WAGI.N)の同じような値に対して、採取されたデータにおいて複数の異なるCPの値が得られていたため、システムはそれらのCPの値の期待値を出力したものと考えられるのである。このような場合には、もしかすると逆に予測値を大きく下回る観測値が得られていた可能性があったといえる。

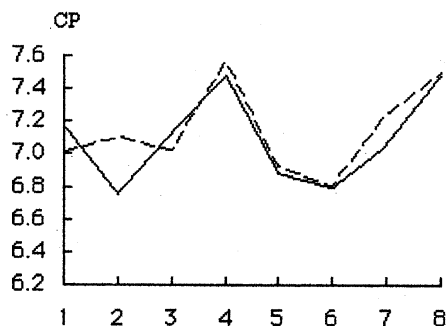


図5(a) 一括して予測した場合

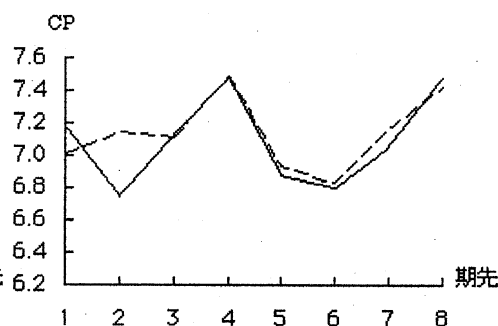


図5(b) 次の期ごとに予測した場合

5. まとめ

本研究では、採取されたデータの要因間の影響分析および操作可能な説明要因の値決定支援のためのシステムを構築した。データとして特に日本経済のマクロデータの一部を取上げ、操作不可能な説明要因として消費者物価指数全国総平均、名目国内総生産ならびに全産業の名目賃金指数、操作可能な説明要因として長期プライムレートを考え、被説明変数として、実質民間最終消費支出を考えた。

提案したシステムはニューラルネットワークを組込んだことで非線形モデルに対応できモデル化において誤差を小さくできる。また、学習後の結果を用いて各説明要因が被説明変数にどのように影響しているのかを示すことができた。さらに、逆問題を解く能力を備えたニューラルネットワークを採用しているため、操作不可能な説明要因を与えることで、操作可能な説明要因である長期プライムレートをどのように操作すれば望ましい実質民間最終消費支出が期待できるのかを示すことができることを示した。

参考文献

- [1] D. E. Rumelhart, J. L. McClelland and the PDP Research Group, "Parallel Distributed Processing," I, MIT Press, 1986.
- [2] R. L. Streitt and T. E. Luginbuhl, "Maximum likelihood training of probabilistic neural networks," *IEEE Trans. NN*, 5, No. 3, pp. 764-783, 1994.
- [3] 福水健次, 渡辺澄夫, "統計的推論を実現するニューラルネットワークとそのパターン認識への応用" 信学技報, NC92-36, pp. 83-90, 1992.
- [4] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc. Ser. B*, 39, pp. 1-38, 1977.