may.13/2022

Neural Network Modelによる人口推計

目白大学人間学部　准教授
村田 久

1.はじめに

Neural Network Model（以下：ニューラルネットワーク）は生体の神経細胞網（脳）をヒントに得て生まれた情報処理モデルであり、その特徴を一言であらわすと非線形動作にあるといわれている。ニューラルネットワークは脳のように外部からの入力に反応して学習、適応する能力構造を持つ。データをニューラルネットワークに与えると、ニューラルネットワークは未知であったデータ構造の関係性を把握し、複雑な非線形写像を表現できることが知られている。

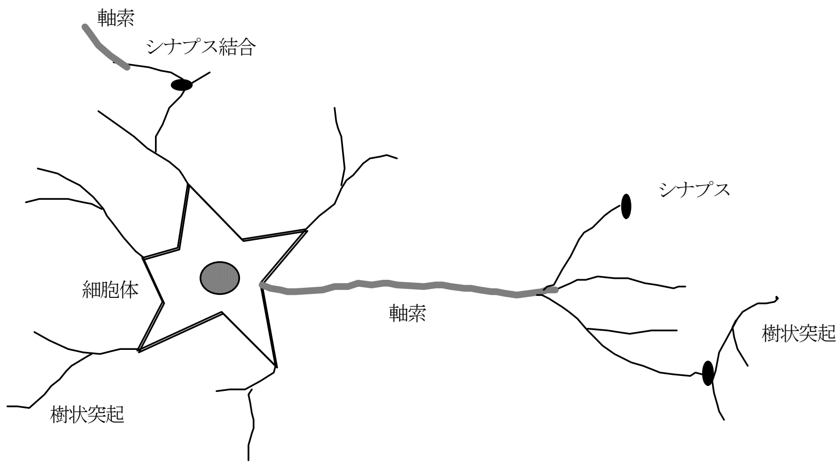
一方人口統計学的パラメータの時系列推移は一般的に非線形であることが知られている。非線形現象は一義的な手法で解析することが困難で、様々な非線形関数が提案されてきた。ニューラルネットワークはその構造から柔軟に関数を構築することが可能でデータによる制約を受けないことで知られている。また様々な領域（株価、為替、気象予報、水量予測・・・）で実際に用いられており、その有用性が報告されている。本稿では、ニューラルネットワークを用いた人口統計学的パラメータの近似及推計の方法と理論について実際の適用例を示しながらそのパフォーマンスについて検討していく。

人口推計の基本的な方法については、大きくは①数学的方法、②要因別推計法の二つに分けることができる。①は既存の人口データの時系列推移から変動の傾向を知る方法であり、数学的曲線が当てはめられる。②は人口動態統計等から要因別に変動を把握する方法で、外生的要因を用いて推計する。本稿では、前者の数学的方法による推計方法をとりあげる。

2.ニューラルネットワークモデル

ニューラルネットワークは生体の神経細胞網を数理モデルとして表現することにより構築される。神経細胞網は個々のニューロン（neuron）の集合体であり、1つのニューロンは細胞体、樹状突起、軸策の3つの部分から構成される（図表 1）。それぞれの役割については、樹状突起は他のニューロンから情報を受け取る入力部であり、その名の通り数十本に枝分かれしてのびている。軸策は出力部であり、軸策の末端にはシナプスがあり、ここから他のニューロンに情報が伝達される。この軸策と樹状突起の結合はシナプス結合とよばれる。軸策に枝分かれはなく1個のニューロンに1つの軸策しか存在しない。細胞体は演算部の役割をつかさどっている。

図表 1　ニューロンの模式図



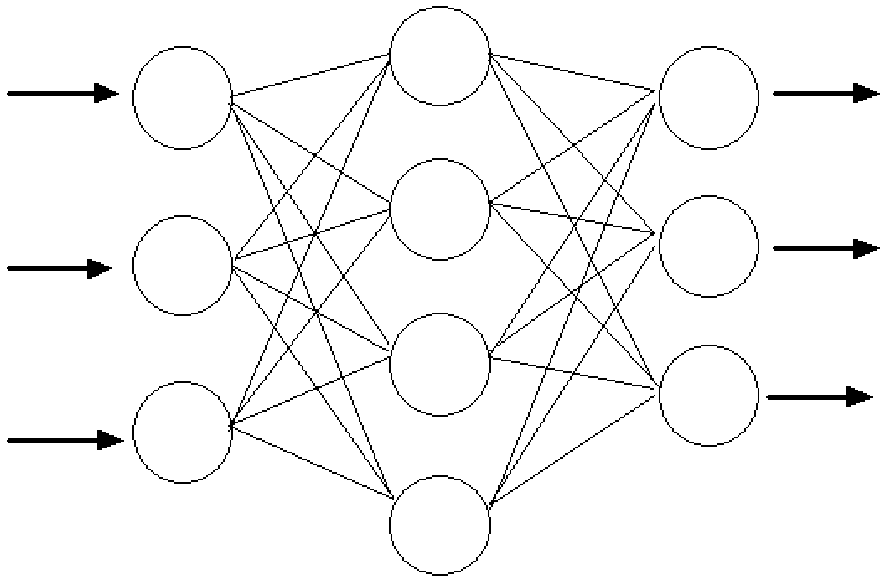
単一のニューロンの数理モデルは次式であらわされる。

U=Σwixj　−　θ
y=ｆ(u)

　ニューロンは重み付け加算と閾値処理の機能を有することが知らされており、wi は加重、θは閾値、xj は入力、y は出力を示している。ニューロンの出力関数 f (u)にはシグモイド関数を用いる。1 つのニューロンが複数の他のニューロンから受け取った刺激の総和が、そのニューロンごとに決められた閾値を超えると興奮し、他の結合しているニューロンに刺激を送る。

　ニューロンをいくつか結合することによりニューラルネットワークのモデルができあがる。ニューラルネットワークには様々な形態があるがここでは階層型ネットワークモデルを採用する。階層型ネットワークモデルは入力層と複数の中間層、および出力層からなり、情報は入力層から出力層に向かって一方向に伝達される（図表 2）。

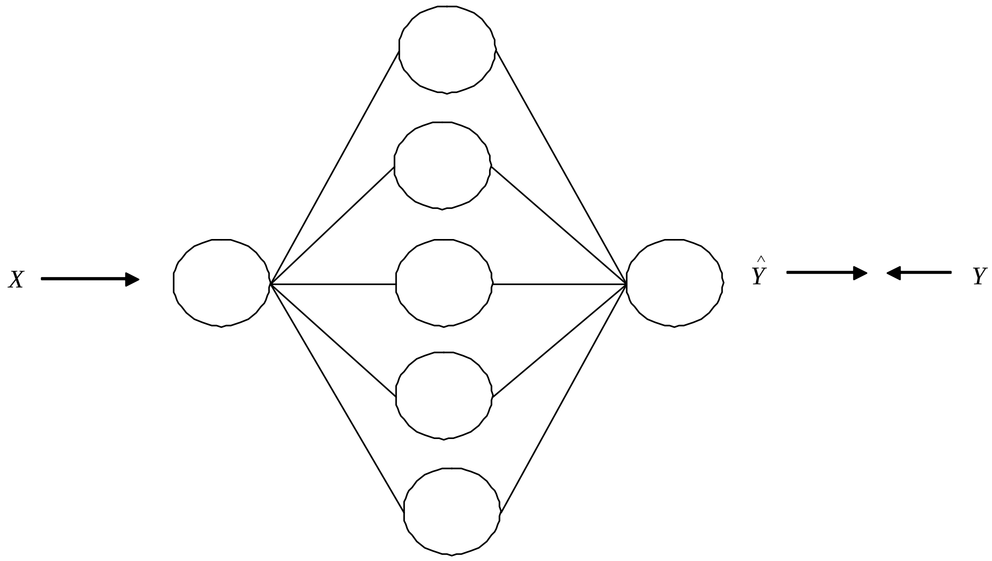
図表 2　階層型ネットワークモデル



入力 xj に対して wi で示された重みを適切に変化させる方法としてバックプロパケーション（誤差逆伝播法）と呼ばれるアルゴリズムを用いる。ニューラルネットワークが出力した答えと正しい答え（教師信号）を比較し、その誤差を最適な重みへの変化量の基準として利用するものである。

ここで人口統計学変数の推計に用いられるニューラルネットワークモデルは図表 3 で示す 3 層モデルである。入力層と出力層の 2 層のみのネットワークでは非線形動作に限界がり、中間層に特徴検知能力があることが知られている。中間層のニューロンの数は 5 つに設定した。入力には通常の回帰式によるものと同様に、基準年を X＝1 とし、翌年（基準年の翌年）を X＝2、さらに翌年を X＝3、・・・・として設定し、教師信号 Y にはその年の人口統計学的パラメータを用いる。

図表 3　推計用ニューラルネットワークモデル



3.シミュレーションの結果

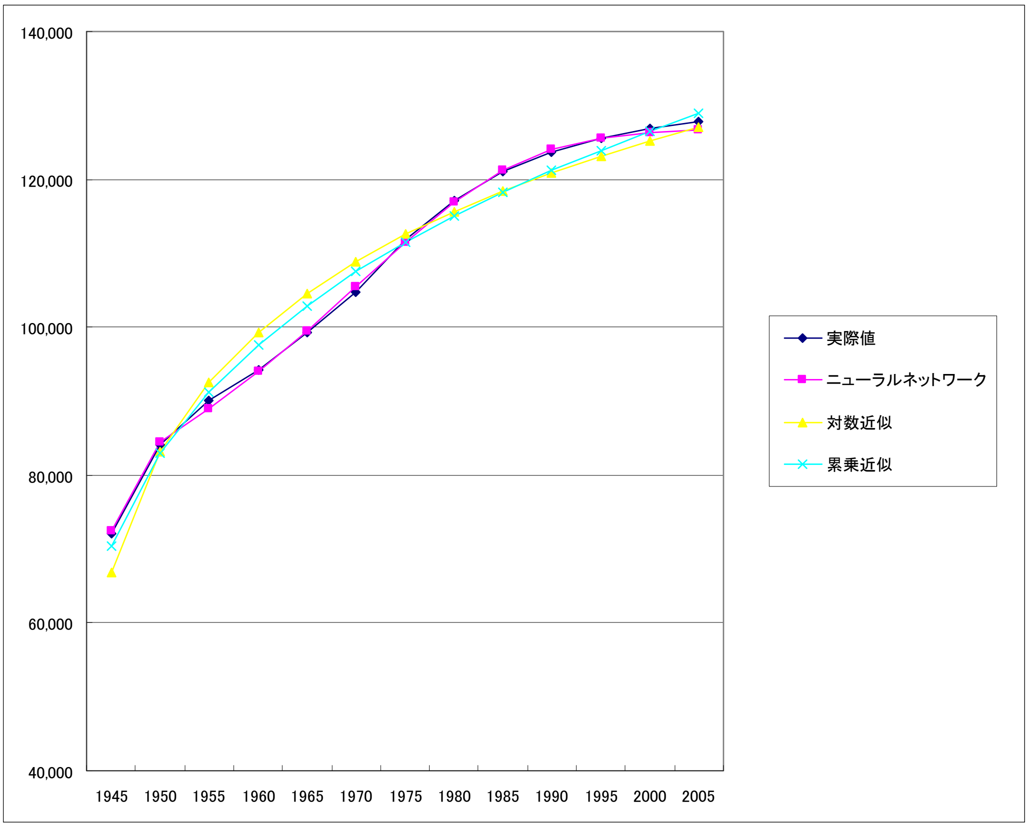
（1）人口

①全国

1945－2005 年期間の日本の全国人口データによる内挿のシミュレーションを行う。ニューラルネットワークの比較対象としてとりあげる近似曲線は対数近似[1]と累乗近似[1]である。推計の結果は図表 4 に示した通りである。

ニューラルネットワークによる結果は実際値との誤差はほとんど見られないが、対数近似及び累乗近似による結果は実際値と乖離が見られ、1955 年－1970 の区間においては上方にズレが見られ、1980 年－2000 年の区間においては下方にズレが見られる。それぞれの期間におけるトレンドを上手く近似できていないことわかる。また、決定係数をみても、ニューラルネットワークは最も高く、0.999 となっており、対数近似では 0.966、累乗近似では 0.985 となっている。

図表 4　日本全国人口シミュレーション結果



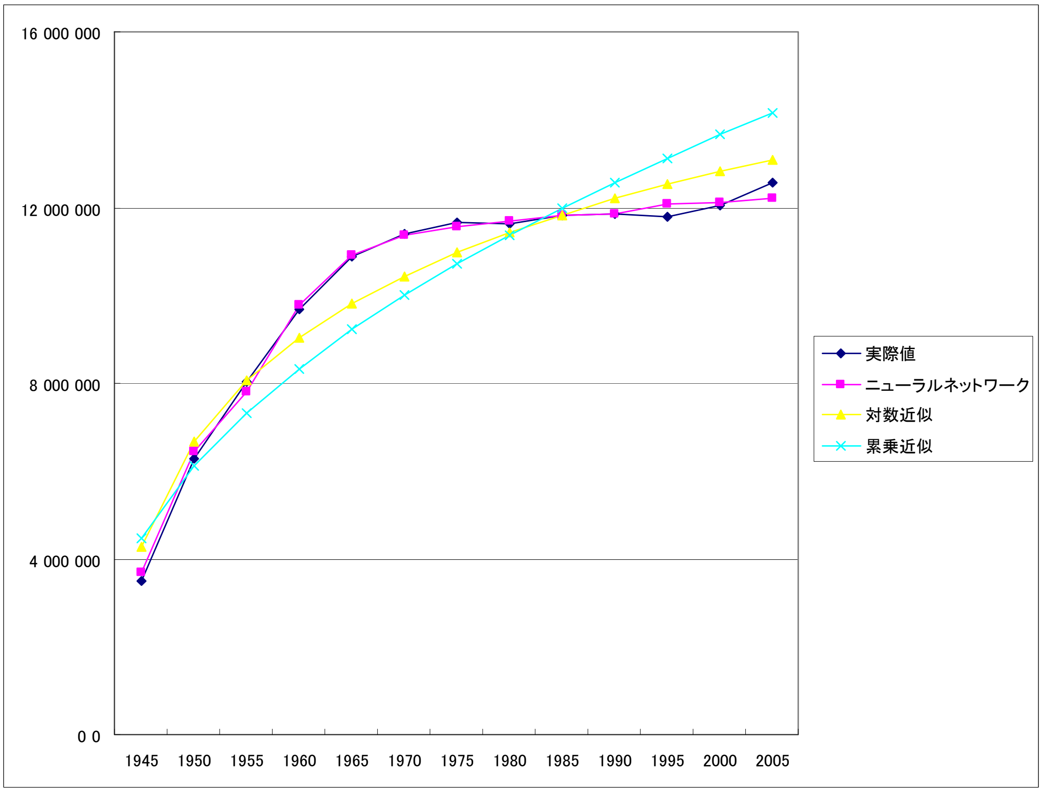
[1] 対数関数　$y = a\log_e x + b$

[1] 累乗の式　$y = ax^b$

②東京都

1945－2005 年期間の東京都人口データによる内挿によるシミュレーションの結果は図表 5 に示した通りである。ニューラルネットワークによる結果は実際値と誤差はほとんど見られないが、対数近似及び累乗近似よる結果では乖離が見られる。1955 年－1980 年の区間においては下方に、1990 年－2005 年においては上方にズレがあることがわかる。決定係数をみても、ニューラルネットワークは 0.996、対数近似は 0.940、累乗近似は 0.850 であり、ニューラルネットワークは最も良いパフォーマンスを示しているのが明らかである。

図表 5　東京都人口シミュレーション結果

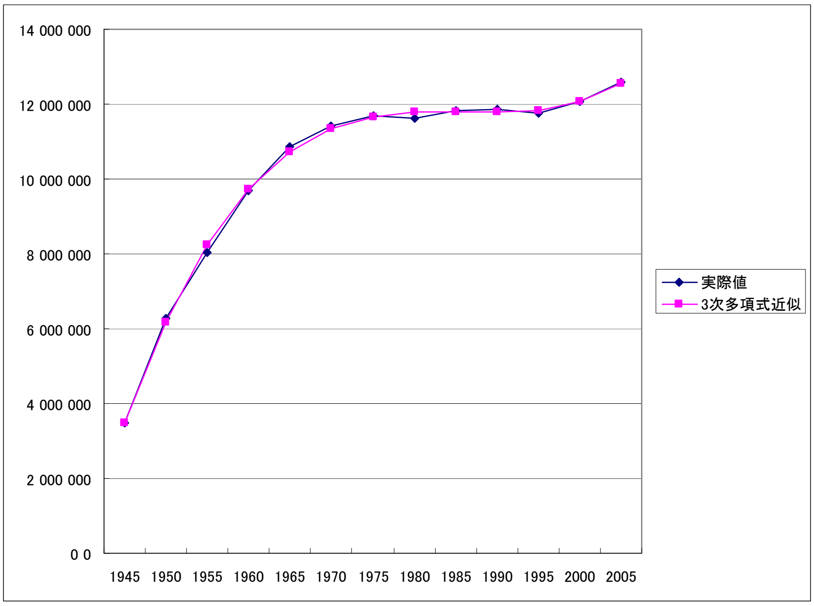


東京都の人口推移は戦後から 1970 年の高度経済成長期には急速に増加し、その後 1975 年から 1990 年代までは安定し、その後 10 年くらいは微増の増加傾向が見られる。このような推移傾向では N 字型の 3 次多項式で近似させるのが適当と判断できる。3 次多項式による近似の結果は図表　にあらわした通りである。

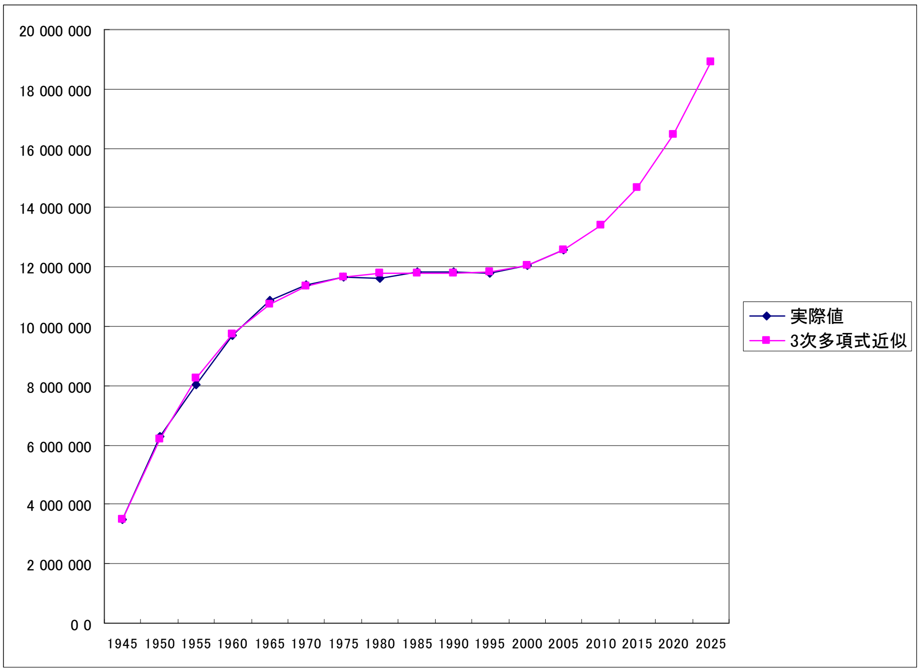
図表 6 を見ると 3 次多項式ではよく近似していることがわかる。決定係数も 0.998 となっている。この関数を使用して 2005 年時点で 20 年後の将来人口を推計した結果を表したものが図表 7 である。3 次多項式の数学的性質上、将来の年次推移により急速に人口が増加する推計となってしまい、オーバーフィットの問題が起こっていることがわかる。現在東京都は都心回帰現象による微増の推移が見られるが、全国

でもっとも低い合計特殊出生率でありこのように急激に単調増加することは考えられず、将来人口の推計には適切ではないことがわかる。

図表６　３次多項式による近似



図表７　３次多項式による将来推計



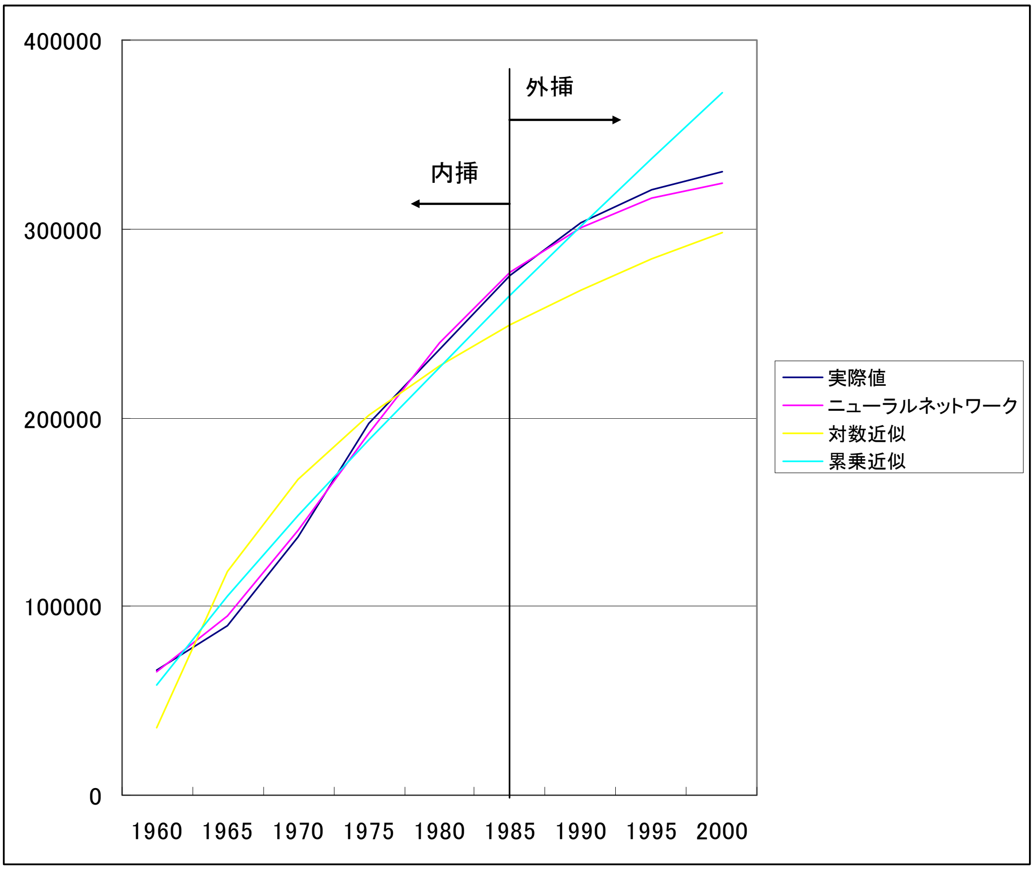
③所沢市

ここでは埼玉県所沢市の 1960 年－2000 年期間人口データを使用し、内挿及び外挿のシミュレーションを行う。内層に用いるデータは 1960－85 年までの 6 時点における人口データであり、1990－2000 年までのデータは外挿の評価に用いる。

図表 8 は 1960 年－2000 年の内挿及び外挿の結果を示したものである。内挿をみると、ニューラルネットワークによる結果は実際値を良く近似しているのに対し、対数近似は近似が悪く、累乗近似においてもニューラルネットワークよりズレが多いことがみてとれる。決定係数でみると、ニューラルネットワークは 0.999、対数近似は 0.901、累乗近似は 0.960 となっておりニューラルネットワークは他と比較し高い値を示している。

外挿でみると、ニューラルネットワークは実際値を良く推計しているのに対し、対数近似による推計は大きく下方にズレており、累乗近似による推計では 1995 年、2000 年において上方にズレが大きいことがみてとれる。

図表 8 所沢市人口シミュレーション結果



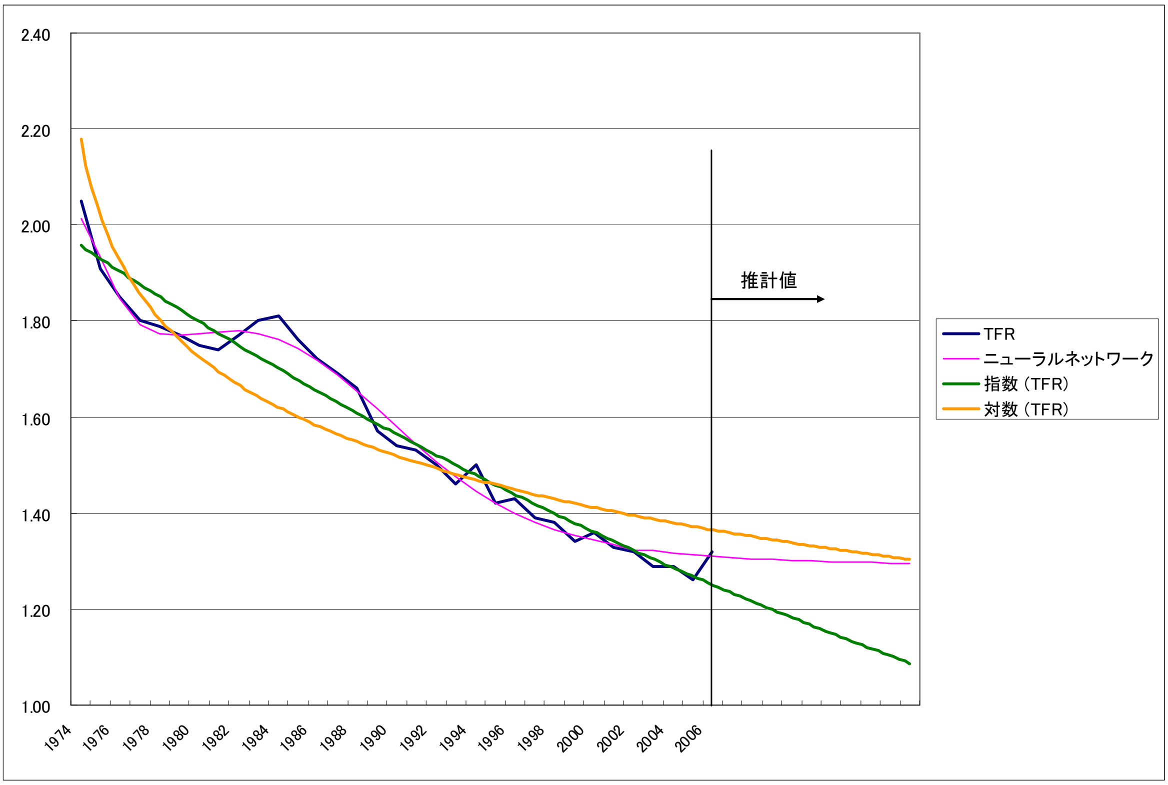
（2）出生率

1974 年－2006 年期間の合計特殊出生率データによるシミュレーションを行う。ニューラルネットワークと比較する近似曲線は指数近似 と対数近似である。図表 9 は 1974 年－2006 年の内挿と 2007 年から 10 年にわたる外挿の結果を表したものである。

出生率の推移変化を最もよく近似しているのはニューラルネットワークであることがみてとれる。1982 年－1984 年区間の上昇トレンドも含みながらなだらかに近似していることがわかる。決定係数を見ても、ニューラルネットワークは 0.987、指数近似は 0.656、対数近似は 0.839 であり、ニューラルネットワークの値が最も良かった。

外挿の結果を見ると、ニューラルネットワークでは 1.3 付近で下げ止まっているのに対し、指数近似では 1.2 を割り込む単調減少を推計し、対数近似では 1996 年以降の内挿で実際より上方に大きくズレた近似となっており、その影響のため 2007 年以降の外挿でも上方の推計値となっていることがみてとれる。

図表 9 出生率シミュレーション結果



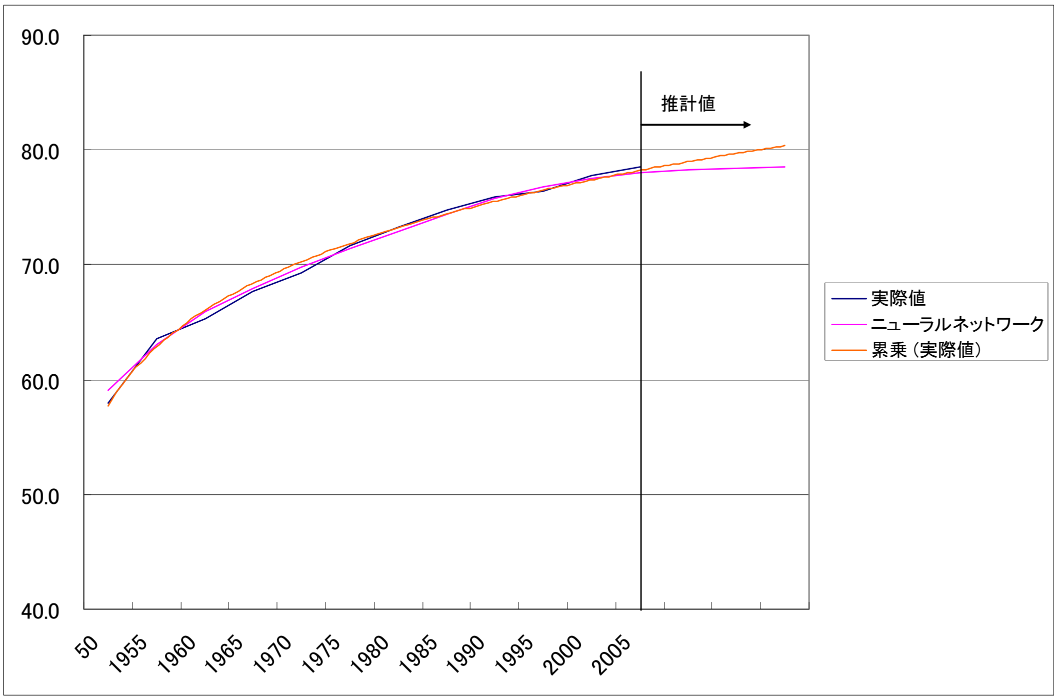
指数関数 y=aebx + c

1950 年－2005 年期間の男子平均寿命データによるシミュレーションを行う。ニューラルネットワークと比較する近似曲線は累乗近似である。図表 10 は 1950 年－2005 年の内挿と 2010 年、2015 年、2020 年の外挿の結果を表したものである。

内挿では平均寿命の実際値の推移はなだらかな単調増加であり、視覚的にはニューラルネットワークと累乗近似に大きな差異は見られない。決定係数についてはニューラルネットワークは 0.995、累乗近似は 0.993 となっている。

外挿の結果を見ると、累乗近似では 2020 年には男子の平均寿命は 80 歳を超えると推計し、ニューラルネットワークでは 79 歳程度と推計している。

図表 10　平均寿命シミュレーション結果



4.まとめ

本稿で示したシミュレーションの結果では、人口、出生率、平均寿命の人口学的変数の近似及び推計において、従来用いられてきた非線形曲線による当てはめの結果よりニューラルネットワークの結果は同等以上のパフォーマンスが得られ、その有用性が示されたといえる。

データが非線形な性質を有する時、その構造を最もよく近似できる予測式は一義的に決定することはできない。予測式の状態はデータの構造に依存し、どのような非線形回帰式を用いるかは分析者に委ねられる。しかし、ニューラルネットワークによる人口推計では所与のデータから関数を学習するので、先見的な予測式を必要としない。つまりどのような人口学的変数に対しても同様のニューラルネットワークモデルを適用することができる。

変数変換後の線形関数または高次回帰モデルなど母数に関して線形な予測式は入力の微小な相違が出力に大きな相違を与える可能性が高く、本稿の東京都人口データで示したように内挿で近似のパフォーマンスが良くても外挿データに対して十分な安定性が保障されない。ニューラルネットワークにおいては内挿で十分な近似が得られるとともに、外挿に対しても合理的な推計結果が得られたと考える。

ニューラルネットワークによる推計が従来方法と比較して同等以上のパフォーマンスが示されたとはいえ、これによりニューラルネットワークによる方法が従来方法にとって変わるということではなく、今後は並列的に使用し、推計方法のバリエーションの 1 つに加えていく方向性が一つとして考えられる。

ニューラルネットワークを利用した推計システムは実際の産業、製品、経済領域で広く応用されているが、推計のメカニズムについてはブラックボックス化しており、ニューラルネットワークによる推計の利点の理論的解明が今後の課題である。また、これらの課題解明はより精度の高い推計に貢献していくと考える。

参考文献

石川晃（1993）『市町村人口推計マニュアル』古今書院

市川紘（1993）『階層型ニューラルネットワーク―非線形問題解析への応用』共立出版

豊田秀樹（1996）『非線形多変量解析―ニューラルネットによるアプローチ』朝倉書店

村田久（2004）ニューラルネットワークによる統計分析の意義．ESTRELA 128 66-70．

村田久（2005）ニューラルネットワークによる統計分析--非線形解析への応用(6)ニューラルネットワークによる回帰分析．Estrela (133) 68-73.