

複数時系列の相互関係を考慮した長期予測に関する基礎的研究

Long-term Prediction considering the mutual relation among several time series

古田 均*・野村泰稔**
Hitoshi FURUTA, Yasutoshi NOMURA

*工博 関西大学教授 総合情報学部総合情報学科(〒569-1095 大阪府高槻市霊仙寺町 2-1-1)

**工修 関西大学大学院 総合情報学研究科総合情報学専攻

Time series analysis is one of important issues in science, meteorology, engineering, and so on. Up to the present, various methods, for instance, Statistical methods, Neural Network, Chaos theory and GMDH have been proposed for the prediction of time series data.

However, Neural Network needs a great deal of learning data and Chaos theory also needs a great deal of time series data when reconstructing the attractor in the state space.

In this study, we propose a long-term prediction method of time series data using Artificial Life considering the Mutual Relation among several time series.

Key Words: Times series analysis, Long-term Prediction, Short-term Prediction, Artificial Life

1. はじめに

時系列予測は工学, 気象学, 経済学など多くの分野で重要課題であり, 予測精度を向上させるために, 統計的手法¹⁾やニューラルネットワーク²⁾, カオス時系列解析法³⁾, GMDH⁴⁾などのソフトコンピューティング技術⁵⁾を用いたものなど, さまざまな手法が提案されてきた。従来は時系列の短期予測を行う際, 統計的解析法が用いられてきた。これは与えられた時系列に対し, その変動を表現する時系列モデルを推定し, 時系列の確率的構造を解析するモデル解析を行い, そして時系列自身が互いに相関を持つことを利用して, 現在までに得られた情報から将来変動の予測を行う。具体的には赤池の方法を用いて自己回帰モデルの項数と係数を求めてカルマンフィルタのエリアを決定し, カルマンフィルタによって自己回帰モデルのパラメータの更新を行いながら予測するという確率的な線形モデルに基づく予測手法が主に用いられている。しかし, 対象とする時系列の挙動の複雑さが確率的要因によるものではなく時系列を生成する系の非線形性に起因する場合は, その系に内在する変数の数, 線形性・非線形性が不明であるので, 必ずしも線形モデルが精度の高い予測結果を与えとは限らない。なおかつさまざまなパラメータを決定しなければならず, パラメータ解析結果が予測精度に大きく影響するという問題点がある。また非線形関数近似に優れた階層型ニューラルネットワークに関しても, 大量の学習データが必要であること, また予測精度はネットワーク各層のニューロン数や時系列そのものの質などに大きく依存してしまうこと, そしてパラメータのチューニングは試行錯誤で決定していくため非常に手間がかかることなど, 若干の問題点がある。

そこで本論文では, 萩原将文氏が提案した人工生命的手法⁶⁾を利用し, 観測されている時系列が少量でかつ, 時系列が複数個存在する場合における複数時系列の同時予測手法を試みる。人工生命技術⁷⁾を用いることにより, 動的に変化する複数時系列の予測を行う。ここでいう人工生命技術とは, 下位層の局所的自律運動の結果, 上位層の大局的秩序が形成されるという意味である。人工生命技術を用いた工学的研究としては機械学習や最適化, またセルオートマトンなどのシミュレーション技術^{8) 9)}への応用を試みる研究が活発に行われている。そしてこれらの試みは有益な結果を残している。このことは人工生命アプローチが従来の方法とは異なり, 局所的な相互作用と大局的な流れを考慮し, さらにフィードバックを有することが大きな要因だと考えられている。

本研究で試みる予測手法は各々の時系列データの1期差分を過去の経験情報として個体に埋め込む。そして各個体同士が相互作用を行った結果を大域的にまとめることで各時系列の予測を行っていく。本論文では, まず2章で, 本研究で提案した手法について説明する。ちなみに本論文では, 萩原氏が提案した手法の説明はしていない。3章では提案手法を実際の9つの時系列を用いて多変数同時予測を行った結果を示す。そして4章で本研究をまとめる。

2. 複数時系列の相互関係を考慮した予測手法

本章では, 本研究で試みた人工生命技術を用いた複数時系列の予測法について述べる。本研究での個体はある時系列の1時点のデータと目標値を保持し, 自分が保持するデータが目標に対してどれくらい活性であるかを表す活性値と, その時点において, 各々の時系列の1期差分を過去

の経験情報として保持する。そしてそれぞれの個体が人工生命フィールド上で局所的に互いのデータを参考にしながら、自分のデータを更新していく。次に、人工生命フィールド上で得られた局所的な作用を大局的なルールとして抽出する。具体的には各個体の保持するデータを平均化する。これを1期先の予測値とする。

2. 1 個体の生成

個体は各々の時系列の数値データから生成される。また各時系列に関して複数の個体が存在する。ここで時刻 t における時系列 i の値を $D_i(t)$ 、その正規化値を $C_i(t)$ とする。個体 X へのコーディングは時系列 i の値が $C_i(t)$ であった時、他の時系列 j ($j=1,2,\dots,n$) は $\Delta C_j(t)$ 増加したという過去の経験情報を差分データベクトルとして、個体にそのまま当てはめる。

$$C_i(t) = \frac{D_i(t) - \min_t(D_i(t))}{\max_t(D_i(t)) - \min_t(D_i(t))} \quad (1)$$

$$\Delta C_j(t) = C_j(t) - C_j(t-1) \quad (2)$$

であり、ここで式-1は時系列の正規化、式-2は経験情報である。それぞれの経験情報は1つの時系列に t 件の数値データが観測されていれば $(t-1)$ 個になる。また個体数は時系列数 $\times (t-1)$ である。

そして複数の時系列の相互相関関数、自己相関関数をあらかじめ求めておく、これは個体間のデータの更新の時に使用する。そして活性値の初期値はあらかじめ予測をする最初の時刻のデータとする。この手順で生成された個体群を2次元の人工生命フィールド上にランダムにばらまく。各個体は8方向にランダムに移動する。

2. 2 個体間の局所的な作用

各個体は相互作用により活性値、経験情報を更新していく。その更新には、局所的な相互作用によるものと、大域的な相互作用によるものがある。ここでは、まず局所的な相互作用について簡単に説明する。ここで、時系列1から生成された個体 A が時系列2から生成された個体 B に近づいたとする。以下に計算方法を示す。

(1) 活性値の計算

人工生命フィールド上で移動する個体同士が隣り合った場合、近づいていった個体 A が近づかれた相手である個体 B の活性値を更新する。ここで、個体 A のルール適合度 R_A をはじめに計算する。ルール適合度とは自分の保持するデータ A_A が活性値である目標値 P_A にどれだけ近いのかを示す指標である。下の式-3に示す。

$$R_A = \exp(-\alpha(P_A - A_A)^2) \quad (3)$$

ここで α はルール適合度を決定する際の定数であり、この

値が大きければ大きいほど、ルールの適合判定が厳しくなる。次に先ほど求めた個体 A のルール適合度 R_A を用いて個体 B の活性値を更新する。下の式-4に示す。

$$P_B^{new} = P_B + \beta \cdot R_A \cdot E_A(B) \quad (4)$$

ここで $E_A(B)$ は個体 A が保持する過去の個体 B の経験情報であり、 β は更新幅を決定する定数である。

(2) 経験情報の更新

活性値を受けた側の個体 B は個体 A の保持するルールを参考に自分の経験情報を更新する。ここでは、あらかじめ求めておいた各々の相互相関関数、またそれぞれの個体が同じ時系列から生成されたもの同士なら自己相関関数により、データの更新が行われる。またもとの時系列のどの部分から生成されたのかという時刻の差、つまり遅れ時間を考慮し、個体 A が保持する自分自身の経験情報を $E_A(A)$ を用いて、個体 B が保持する個体 A についての経験情報 $E_B(A)$ の更新が行われる。下の式-5に示す。

$$E_B^{new}(A) = E_B(A) + R_A \cdot \text{Corr}_{(A,B)} |t_A - t_B| \cdot E_A(A) \quad (5)$$

以上2つの手順を全ての個体に関してランダムに繰り返す。そして全ての個体が移動を終えた時を1ステップとする。

2. 3 予測値の算出

次に個体間の局所的な相互作用の結果を予測値として算出する。また本研究では、ルール生成の際に用いたデータと個体の動作の同期をとるためにある一定のステップを1世代として考える。この際のステップ数は試行錯誤によって決定した。本研究で試みた手法では、各々の時系列から生成される個体群がある時刻において、局所的な相互作用から得られる活性値がどれくらい活発になっていくかを推定しようとしている。そこで同じ時系列から生成された個体を群として評価しなければならない。そして同じ時系列から生成された個体が保持する活性値の平均化を行うことにより、提案手法の予測値とする。時系列 i から生成された個体の数を m_i とすると時系列 i の世代 t における値 C_i^t は下の式-6のように求められる。

$$C_i^t = \sum_A^{m_i} P_A / m_i \quad (6)$$

またそれぞれの各個体の保持する経験情報も同じように平均化し、群全体のルールとする。また短期予測と長期予測で若干処理手順が変わるので下に記す。

(1) 短期予測の処理手順

各々の時系列から、ある一定時刻まで生成された個体を学習データと呼ぶ。各個体を人工生命フィールド上にラン

ダムにばらまき、上述してきた処理手順で局所的な相互作用を行い、個体が保持する情報を更新させる。そしてある一定ステップで全てのデータを平均化することによって、各々の時系列においての1期先の予測値とする。そして次に、各個体を各々の時系列による2件目から先ほどの予測値と対である観測されるデータを用いて、個体の再生成をし、上述の手順を繰り返す。というように、本研究で試みる短期予測においては、平均化された、または予測されたもろもろの情報をフィードバックしない。つまり、局所的な相互作用だけによる予測を行う。このような手順で1期先の予測を数回繰り返している。

(2) 長期予測の処理手順

長期予測においては、活性値の平均化、またすべての経験情報を平均化し、これを大域的なルールとする。そしてその平均化することによって得られた大域的なルールを人工生命フィールド上にフィードバックする。具体的には、特に交渉回数の少なかった個体やまた交渉回数が多すぎて、偏ったデータを持っているデータをリセットし、平均化されたデータをフィードバックすることにより、すべての個体が環境から受けた影響を考慮し、個体を書き換えることになる。本研究では1期先で観測される実測値を読み込まず、各々の予測値を各個体にフィードバックしていくことで長期予測を実現している。

3. 数値計算シミュレーション

本研究では、以下に示すような日本統計年鑑中の記載されている総務庁統計局、経済企画庁からの数値データ¹⁰⁾、1954年から1993年の40年間分を使用した。その理由としてまず、完全失業者の増加問題は以前から大きな社会問題であり、また完全失業者の増加の推移の把握は、あらゆる問題との密接な関連性があり、大変重要であることなどがあげられる。

実際に使用した時系列は以下の9つである。

- D-1：国民総生産における第1次産業の占める割合
- D-2：同第2次産業の占める割合
- D-3：同第3次産業の占める割合
- D-4：総人口
- D-5：労働力人口
- D-6：就業者の産業別割合（第1次産業）
- D-7：同第2次産業
- D-8：同第3次産業
- D-9：完全失業者数

また、使用した時系列の単位は割合と人数である。それぞれの時系列を式-1によって正規化している。またその正規化された結果を図-1に示す。

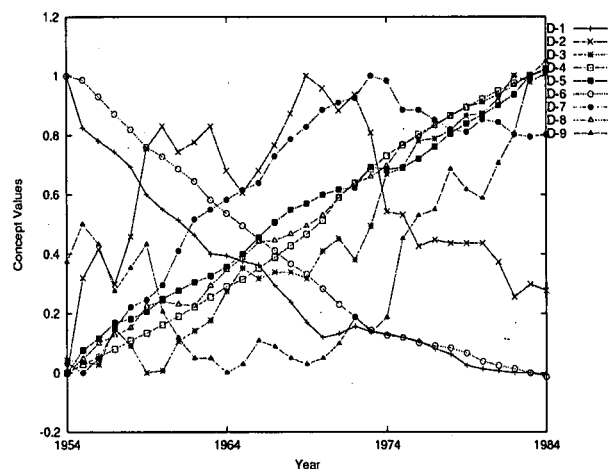


図-1 ルール生成に用いた複数時系列

これらの数値データは図-1のように下降傾向にある時系列、一度上昇して再び下降する時系列、常に上昇傾向にある時系列、また上下に運動する複雑な挙動を示す時系列などが含まれている。

3. 1 提案手法の短期的・長期予測結果

人工生命技術はボトムアップ的なアプローチが主流である。しかし実際にはトップダウン的な作業がなければ、安定した解を得られないこともまれにある。本研究では、あらかじめ1954年から30年間分のデータを使用して時系列の正規化、相互相関関数、自己相関関数、活性値の計算、経験情報の計算、個体生成を行った。また生成された個体の数は $(30年 - 1) \times 9$ 時系列の291個体である。そしてその後の10年間分について短期予測、長期予測を試みた。各個体の活性値の初期値として1984年の値を使用した。またパラメータ設定に関しては α の値を1.0~5.0、人工生命フィールドの広さを500×500とさまざま変化させてみたが、あまり顕著な違いを見出すことはできなかった。また本研究で提案した予測手法は複数の時系列の相互相関関数、自己相関関数などの単純な相互関係、また人工生命フィールド上の単純な局所的な相互作用を考慮している。また現在までに提案されている手法で複数時系列の相互関係を考慮したものは少なく、複数時系列の相互関係を考慮した場合としない場合では、考慮した場合の方が高精度な予測結果を得る可能性が高いと思われる。そして短期予測に関しては、乱数を使用しているにもかかわらず、解の安定性が強く、ほぼ同一の予測結果、また実際の数値データに近い動作を確認することができた。以下の表-1に予測に関するパラメータ設定値を示す。また実測値を下の図-2に、短期予測手法における予測結果を図-3に、長期予測手法による予測結果を図-4に示す。

表-1 各パラメータ設定値

フィールドの広さ	300×300
1世代のステップ数	30
α	3.0
β	1.0
学習データ	30年分
予測世代数	10世代
1世代のステップ数	30回

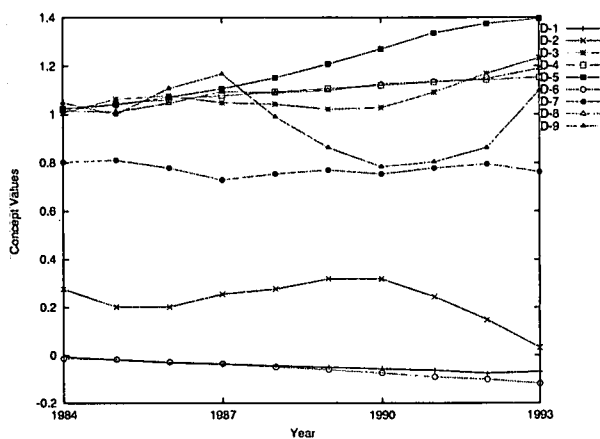


図-2 10年間の実測データ

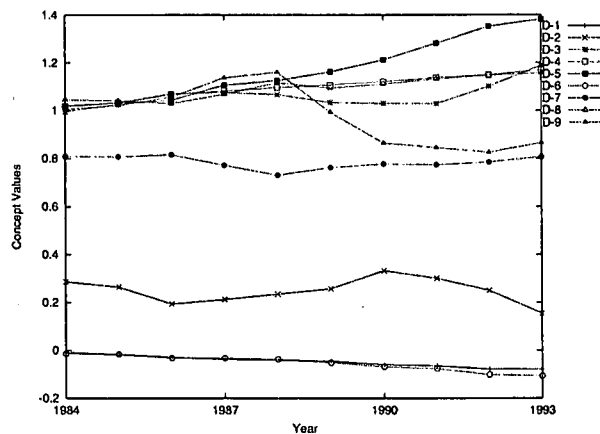


図-3 提案手法による短期予測

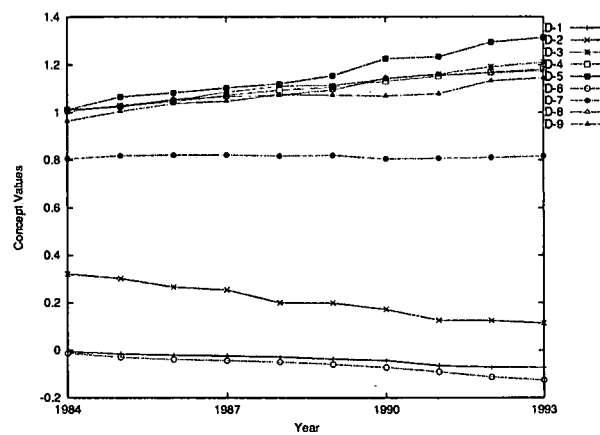


図-4 提案手法による長期予測

短期予測に関しては、複雑な挙動をする時系列も高精度な予測結果を与えることができた。しかし長期予測に関しては、短期予測に比べると、あまり精度の良い結果を得ることはできなかった。

3. 2 短期予測による結果の考察

1つずつ時系列の変動を可視化して予測精度の検証を行う。カオス時系列解析、ニューラルネットワーク理論などでの予測結果で良く見受けられるが、若干、予測値が1世代ずれ、遅れていることがわかる。しかし、常に上昇傾向にある時系列や下降傾向にある時系列だけではなく、複雑に挙動するD-2、D-3、D-7、D-9などの時系列も、かなり高精度に複雑な挙動を再現できていることがわかる。結果を下の図-5～13に示す。

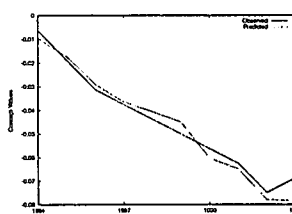


図-5 D-1

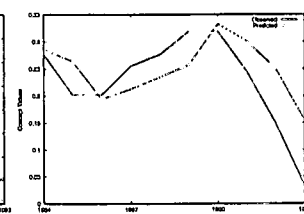


図-6 D-2

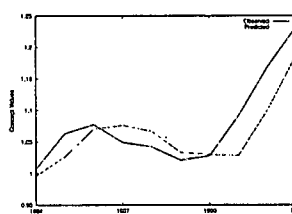


図-7 D-3

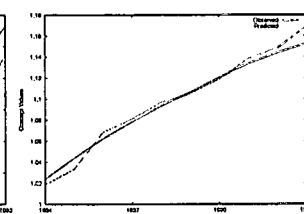


図-8 D-4

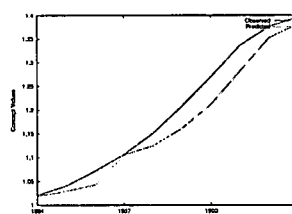


図-9 D-5

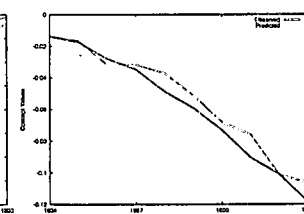


図-10 D-6

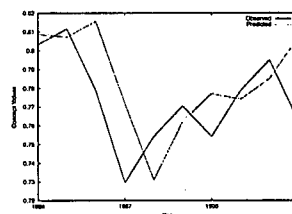


図-11 D-7

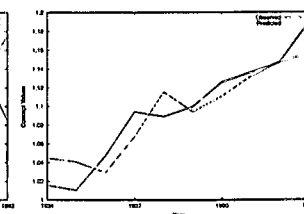


図-12 D-8

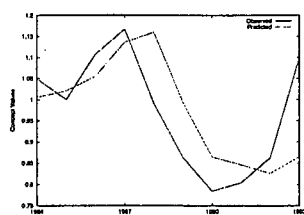


図-1 3 D-9

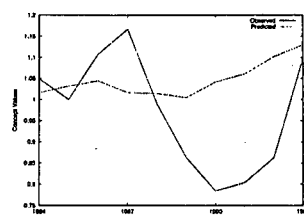


図-2 2 D-9

3. 3 長期予測による結果の考察

次に長期予測結果を1時系列ずつ可視化して検証する。常に上昇傾向にある時系列や下降傾向にある時系列などは短期予測に比べても、顕著な違いを見出すことはできない、しかしD-2, D-3, D-7, D-9などの複雑な振る舞いを見せる時系列に関しては、複雑な挙動の再現性には欠けている。ゆえに精度の良い結果を得ることはできなかったと言える。下の図-1 4～2 2に示す。また表-2に短期的・長期予測の結果を相関係数で示す。

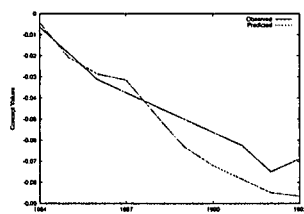


図-1 4 D-1

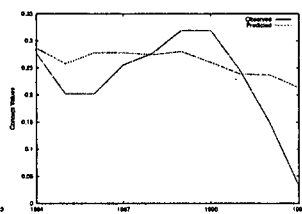


図-1 5 D-2

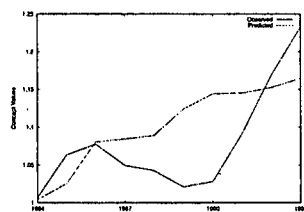


図-1 6 D-3

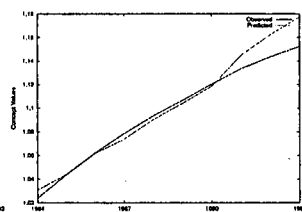


図-1 7 D-4

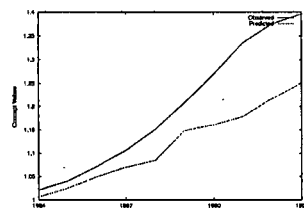


図-1 8 D-5

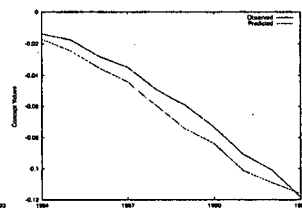


図-1 9 D-6

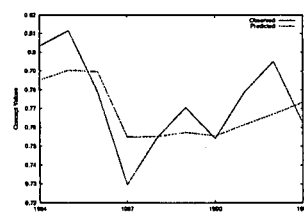


図-2 0 D-7

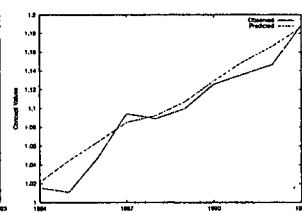


図-2 1 D-8

表-2 相関係数

	Short-term	Long-term
D-1	0.986045651	0.981698883
D-2	0.700864877	0.775920284
D-3	0.880261616	0.588771669
D-4	0.994149922	0.985612531
D-5	0.990285909	0.988499274
D-6	0.988347657	0.991368531
D-7	0.544157615	0.742377872
D-8	0.925286915	0.978650665
D-9	0.619798233	-0.040973056

相関係数からも短期予測の結果が安定しており、0.8以上を9時系列中の6時系列において記録した。しかしグラフからは、時系列の複雑な挙動を追従しているように見えるが、相関係数が低い原因は相関係数を求める系列が10件しかなかったこと、また時系列が複雑な挙動を示す時、1期ずれたことがD-2, D-7, D-9の相関係数の低さに現れていると考えられる。

また長期予測の結果も相関係数からはD-3, D-9以外はかなり高精度な予測結果を得ることができたと言える。

4. おわりに

本研究は萩原将文氏が提案した人工生命的手法を用いた動的に変化する複数概念関係の抽出とその長期予測手法を利用し、さらに複数の時系列の同時短期的・長期予測手法への拡張を試みた。具体的には複数時系列の相互相関性、自己相関性をあらかじめ、算出し各時系列を時間的に分割し、多くの個体群を生成した。そして、人工生命フィールド上で活性値、経験情報を更新しながら、各個体と局所的な相互作用を行い、大局的なルールを抽出しながら、予測を試みた。そして人工生命上にフィードバックする際に相互相関係数、自己相関係数を再計算しながら、次元を進めていった。提案手法はランダム性を含む個体を用いているが、短期予測に関しては解のばらつきも少なく、図-3と同様の結果を見ることができた。また長期予測に関しても、短期予測結果と比べると、若干、解の安定性には欠けるが、ランダム性を含む個体を生成しているとはいえ、全試行の7割ぐらいで、同じような解を見ることができた。そして、同時に複数時系列の短期予測が行えたことがあげられる。また今後の課題として、

さらに

- ・ さらなる予測精度の向上
- ・ さまざまな複数時系列への適用
- ・ また予測段階での新しい時系列の動的な追加方法の検討
- ・ 外的な環境の変化を考慮したより柔軟な推論法の確立

などがあげられる。

参考文献

- 1) N, Sugiura.: Further analysis of the data by Akaike's information criterion and the finite corrections, Comm. In Statist., Vol.A 7, pp.13-16, 1978.
- 2) 熊沢逸夫: 学習とニューラルネットワーク, 電子情報通信工学シリーズ, 森北出版, 1998.
- 3) 五百旗頭正, 菅家正康, 藤本泰成, 鈴木新吾: カオスの時系列の短期予測のための局所ファジィ再構成法,

日本ファジィ学会誌, Vol.7, No.1, pp.186-194, 1995.

- 4) 林勲: GMDH, 日本ファジィ学会誌, Vol.7, No.2, pp.270-274, 1995.
- 5) 坂和正敏, 馬野元秀, 大里有生: ソフトコンピューティング用語集, 朝倉出版, 1996
- 6) 木本武一郎, 萩原将文: 人工生命的手法を用いた動的に変化する複数概念関係の抽出とその長期予測, 日本ファジィ学会誌, Vol.7, No.6, pp.1033-1040, 1999
- 7) 星野力: 人工生命の夢と悩み-コンピュータの中の知能と行動の進化-, 裳華房, 1994
- 8) 加藤恭義, 光成友孝, 築山洋: セルオートマトン法-複雑系の自己組織化と超並列処理-, 森北出版, 1999
- 9) 上田完次, 下原勝憲, 伊庭斉志: 人工生命の方法 - そのパラダイムと研究最前線-, 日本情報処理開発協会, 1995
- 10) 総務庁統計局編: 第44回日本統計年鑑, 日本統計協会, 1995