

適応的ルール生成システムによる株価時系列予測

池田, 欽一

<https://doi.org/10.15017/3000189>

出版情報 : 経済論究. 101, pp.1-12, 1998-07-28. 九州大学大学院経済学会
バージョン :
権利関係 :

適応的ルール生成システムによる株価時系列予測

池 田 欽 一

まえがき

人工知能，知識工学の分野で条件 (if) ～行動 (then) ルール集合を用いて，推論，意志決定を行なうプロダクションシステムに，比較的新しい確率的探索法である遺伝的アルゴリズム (GA) を組み合わせて，そのプロダクションシステムの新たなルールの生成を自立的に行なわせる自動的プログラミングシステム (クラシファイアシステムと呼ばれる) が研究されてきた [1]，[2]，[9]，[10]。本論文では，その概念を応用した，適応的ルール生成システムの構成を試み，株価時系列の予測シミュレーションを行なった。

クラシファイアシステムは，もともと，上にも書いてある通り，人工知能の分野での効率的で，外部環境に適応する知識ベースを構築するものであるが，本論文では，条件となる部分には過去の株価の変動パターンを与え，行動，結論部分は，条件部のパターンに対する予測値であると仮定をおくことにより，外部環境 (株価市場) に適合した株価変動パターン集合の自動生成システムを考える。

まず，2章ではクラシファイアシステムとはどのようなものであるか，また，その概念を用いてどのような株価予測システム (適応的ルール生成システム) を構築するのかを述べる。

3章では，過去の株価変動とそれに対する予測値のペアで表されるルールを，どのように更新していくかを示す。ルール更新には遺伝的アルゴリズム (GA) を用いる。遺伝的アルゴリズムでは，染色体と呼ばれる遺伝子列によって解候補を表し，適応度と呼ばれる評価によって次世代構成へ寄与する割合をきめ，解候補集団全体の最適値への改良を行なうものである。本システムでは予測ルールを個体集団であるとみなし，予測ルールが外部環境 (株価変動) にどの程度あっているかを適応度として効率的ルール集合を求めることになる。

4章では，本システムを日経平均株価，および，ランダムに選択した企業の株価の予測に適用した場合の誤差についてのシミュレーション分析を行なった。なお，比較のため，ARモデルによる予測誤差についてもまとめている。

2 適応的ルール生成システムによる時系列予測

2.1 クラシファイアシステム

クラシファイアシステムとは，*if～then*形式のルールの集合であり，外部環境からの情報入力

により、ある出力を行なうシステムである。

クラシファイアシステムは、もともと遺伝的アルゴリズムを用いて、機械学習を行なうものであり、遺伝子空間上にコーディングされたプログラムを遺伝的アルゴリズムを用いて外部環境に適したものを自動的に生成するためのシステムである。

クラシファイアシステムが対象としている環境は、(1)ある一定数のルールで規定できるような安定した環境でない、(2)ノイズの多い環境、(3)最終的な目標が明確でない環境、(4)出力値に対する一定の評価基準があり、それによって報酬が与えられる環境を前提としている。

株価変動、投資を考えると一般的に非定常であり、さまざまなノイズをもち、はっきりとした目標値はなく、どれだけの収益があったかという評価を簡単に行なうことができるので、前記の環境の条件を満たしているようである。

クラシファイアシステムは以下の3つの機能から構成されている。

- プロダクションシステム
- 報酬の分配、信頼度の伝搬
- 遺伝的アルゴリズムによる新ルール生成システム

2.2 適応的ルール生成システム

プロダクションシステムとは、人工知能分野において推論、判断を行なうモデルである。プロダクションシステムは、コンディション (if) 部、アクション (then) 部の対で表現されるプロダクションルール、プロダクションルールの適応対象となるデータを記録するワーキングメモリ、ワーキングメモリの内容とプロダクションルールのコンディション部のマッチングを行ない入力情報に適合するルールを検索し、そのルールのアクション部に従って外部環境に情報を出力するインタプリタによって構成される。

クラシファイアシステムにおけるプロダクションシステムは、インタプリタに対応する入力インターフェイス、分類リスト、出力インターフェイス、ワーキングメモリに対応するメッセージリスト、

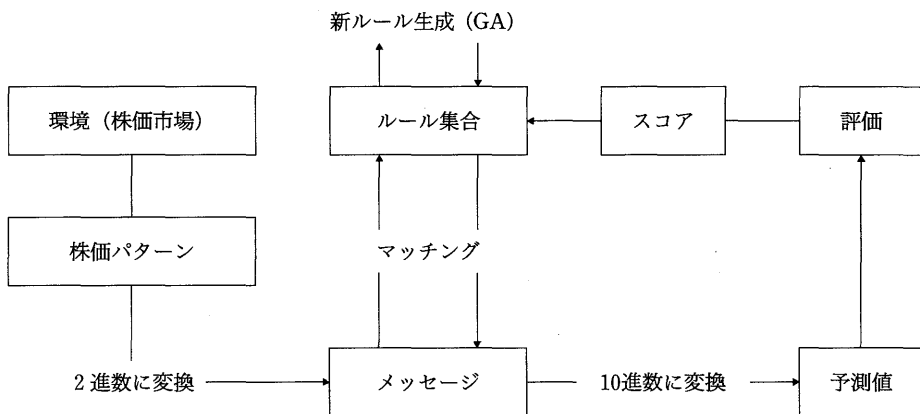


図1 適応的ルール生成システムの構成

そして知識を表すルールからなっている。

メッセージとは外部入力や、内部状態を表す記号列で、アルファベットや、0, 1のバイナリ値によって表現する。適応的ルール生成システムでは、予測に用いる過去の株価変動パターン、それに対する予測値がメッセージに対応しており、システム内部では10進数表記された値を2進数に変換して用いることになる。例えば、0～1の間の値をとるように基準化された株価が0.549となった場合、0, 1によるビット表現（10ビット）は

$$0.549 \rightarrow 0100011001. \tag{1}$$

となる。ルールはメッセージと同様に記号列で表されるコンディション部、アクション部をもち、そのルールの外部環境に対する適応度を表す信頼度という値をもっている。

入力インターフェイスでは、外部からの入力情報を適切な記号列に変換し、メッセージリストに書き込みを行なう。

分類リストは、メッセージリストに書き込まれたメッセージとルールのコンディション部とのマッチングを行ない、条件を満たすルールを検索する。マッチしたルールの内でもっとも信頼度の高いルールが採用され、そのルールのアクション部の記号列が出力インターフェイスによって外部環境への出力とされる。

その後、出力インターフェイスによって出力された値が外部環境に適合しているかどうか評価をおこない、その評価に従って出力を行なったルールに対して報酬が与えられる。

以上のようなサイクルを一定期間繰り返した後、ルールを生物集団、ルールの信頼度を適応度とする遺伝的アルゴリズムによって新たなルール集合を生成してより外部環境に適したルールの集合を構成していく。

クラシファイアシステムを時系列予測に応用した適応的ルール生成システムでは、次の仮定、前提をしておく。ルールの表現方法として、条件部には過去の株価変動パターンを0～1の値に基準化したものを与え、それに対応する結論、行動部には過去の株価変動パターンに対する同様に基準化した

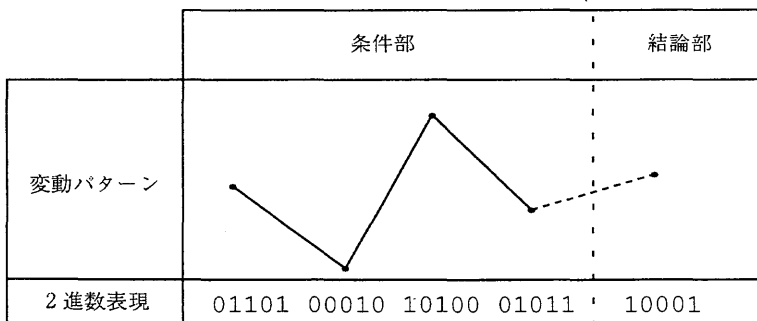


図2 ルールを表現する記号列の例

予測値を与えている（図 2 参照）。

また、ルールには、外部環境に対する信頼値、評価値を与えなければならない。この場合のルールの評価基準としては外部環境より過去の株価変動のパターンが示され、それにマッチした条件部をもつルールが活性化し、その結論部を外部環境に次の期（1日先）の予測値として示すことになる。その予測に対応する実測値が観測された場合、活性化したルールの結論部がどの程度正確な予測値を与えたのか評価できることになる。その評価に対応してルールが外部環境に適しているのかどうかという信頼度を与えるようにすればよい。

以上のような評価を行なうためには、外部環境からメッセージ（過去の株価変動パターン）が送られて来た時、どのルールの条件部とマッチさせるかが重要となってくる。本システムでは個々のルールの条件部と与えられた変動パターンのユークリットの距離がもっとも小さいものをマッチしたと考え、そのルールが活性化されるとした。システムへの入力情報となる株価時系列を $S_t, S_{t+1}, \dots, S_{t+p}$ 、ルール n の条件部の10進数表現列を $P_1^n, P_2^n, \dots, P_p^n$ とすると活性化されるルールは

$$\min_n \sqrt{\sum_{i=1}^p (S_{t+i} - P_i^n)^2}. \quad (2)$$

を満たすルール n となる。

3 遺伝的アルゴリズムによるルールの生成

3.1 遺伝的アルゴリズム (GA)

遺伝的アルゴリズム (GA) とは、生物進化のプロセスを模倣し、問題に対する最適解を広域的に探索する確率的探索アルゴリズムである。

このアルゴリズムの大まかな手順としては、まず解の候補を染色体と呼ばれる遺伝子列で記述した。これはコーディングといわれるもので、GAにおいてもっとも重要な位置をしめている。その後、(1)初期の個体集団をランダムに作成（初期化）し、(2)それぞれの個体の環境への適応度の評価を行ない、(3)その適合度の高いものが次の世代により多く残るように選択する処理をし、(4)さらに交差、(5)突然変異処理などにより新しい世代集団を構成していく。(2)~(5)の手順が繰り返され、さまざまな染色体をもつ個体を進化させもっとも適合する解候補を探索していく。

以下では、それぞれの手順の概要を示す。

3.1.1 コーディング

遺伝的アルゴリズムにおいて、染色体で定義される性質、形質の外部表現が表現型とよばれ、性質、形質の染色体による内部表現を遺伝子型と呼ぶ。例えば、そらまめのDNAによる遺伝子配列表現が遺伝子型であり、その染色体によって実際に形づくられたそらまめ自体が表現型にあたる。表現型から遺伝子型へ変換することをコーディングといい、逆に、遺伝子型から表現型への変換をデコーディングという。染色体を構成する遺伝子記号には、0、1の2値をとるバイナリ型、整数型、実数型、ア

ルファベットなどが考えられる。どの遺伝子記号を用いればよいかは解こうとする最適問題に依存する。

適応的ルール生成システムでは、プロダクションシステムのルールを適切な遺伝子記号列で示す必要がある。ルールの条件部には、過去の株価の変動パターンを与えなければならないので、株価を2進数(0, 1)で表現し、予測に用いる株価の期間分の記号列で表現する。それに対応するルールの結論部には条件部に対する予測値を、同じように2進数で表現した記号列で表す。

p 期過去の値までを予測に用い、株価、予測値を b ビットで表すとすると、1つの染色体のビット数 $bits$ は

$$bits = b(p+1). \quad (3)$$

となる。

3.2 初期化

コーディング過程で、問題に対する解を染色体によってどのように表現するかを決定した後、初期世代集団となる N 個の染色体をランダムに生成する。集団サイズ N は大きい方がより広域的探索が行なえるため、大きい方がよいが計算時間の増大などの問題もあるので、そのことを考慮する必要がある。

予測システムでは $bits$ の長さをもつ記号列を N 個発生させることになる。0~1の一様乱数を $bits \times N$ 個発生させ、0~0.5の範囲の値の場合は1を与え、0.5~1範囲の場合は0を対応させる。

3.3 環境への適応度の評価

ここでは、生成された個体のそれぞれについて環境、つまり最適化したい問題への適応度の計算を行なう。適応度関数を $f(x)$ とし、個体 j の表現型を x_j とすると、

$$fitness_j = f(x_j). \quad (4)$$

と求めることができる。

予測システムでは、環境への適応度の評価を新たに計算する必要はなくそれぞれのルールに与えられた信頼度を環境への適応度とみなすことができる。

信頼度(適応度)が高いルールほど、現在の株価変動特性をよりよく表現していることになる。

3.4 選択

次に、計算された適応度をもとに次世代の個体を生成するための個体を確率的に選択する。適応度の高いものほどより高い確率で選択される。適応度の高い個体の選択確率を高くすれば高くするほど解の収束は速くなる。しかし、そうすると、似たような遺伝子型をもった個体が多くなり、すぐれた部分的な遺伝子型をもっている可能性のある適応度の低い染色体が生き残れない可能性がでてきて、広域的な探索が行なわれない場合がある。逆に、適応度の高いものの選択確率を低くしすぎると、ラ

ングダム探索と変わらない結果となってしまう。

上記のような確率的選択では、もっとも適合した染色体をもった個体を淘汰してしまう可能性があるため、1番適応度の高い個体が残れなかった場合に、無条件にこの染色体を残すエリート保存処理が行なわれることもある。

本システムでは、個々のルールは独立したものではなくルール全体で1つの予測パターンを形成しているため、ルール集合に大きな変更を加えると予測誤差の増大、遺伝的アルゴリズムにおける退化が起こる可能性がある。よって、適応度にあたる信頼度が比較的大きなものは優先して高確率で保存することとした。優先する個体は全体の20%程度とし、その他の個体については信頼度に応じた確率、

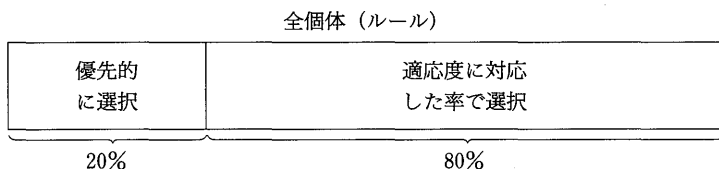


図3 本システムにおける選択

$$p(x_j) = \frac{f(x_j) - \min_i \{f(x_i)\}}{\sum_k (f(x_k) - \min_i \{f(x_i)\})} \tag{5}$$

によって選択を行なう (図3参照)。

3.5 交叉

選択により次世代を構成する個体集団が選ばれた後、その中から2つの個体を親とする新しい子の個体を染色体の組み替えによって作り生成する。交叉方法には染色体の1点のみで交叉を行なう1点交叉法、複数ヶ所で行なう多点交叉法、ある任意のパターンに従い、どちらの親の遺伝子を継承するか決定する一様交叉法、それぞれの親の中間の値を継承させる混合交叉法などがある。どの交叉法を用いるかは、解こうとする最適問題、コーディングの方法に強く依存するので、親の形質が適切に継承されるように選ぶ必要がある。交叉により、選択で生き残った解として適切であろうルールパターンをもった染色体の性質を受け継ぎつつ、他の染色体の特質も受け継いだ新たな染色体ができ、親よりも優れた個体が生成される可能性がある。

予測システムでは、選択された個体を確率 P_{cross} によってランダムに選ばれたペアと多点(2点)で交叉を行なった。

交叉により親ルールの変動パターンを継承しつつ、新たな条件部、結論部をもったルールが生成される。このように新たに生成されたルールは条件部がマッチする外部環境からのメッセージがあった

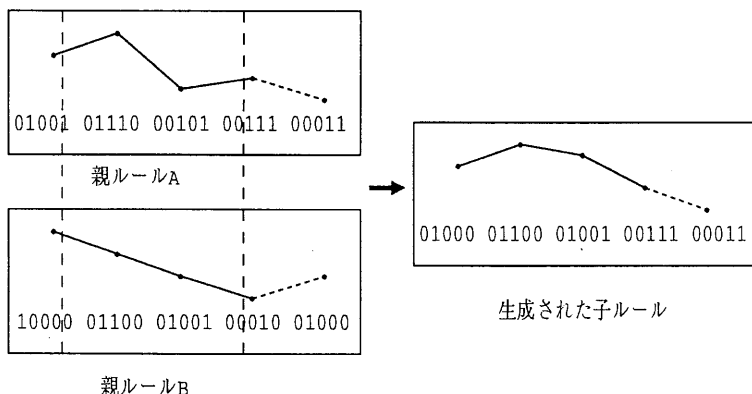


図4 交叉によるルール生成

場合、その結論部を予測値として示し、評価が行なわれ、外部の環境に適している場合は次の世代で選択され、新たなルールを形成していくことになる。図4には、交叉によるルール生成の例を示す。破線で囲まれた遺伝子パターンを親ルールBから継承し、それ以外の遺伝子を親ルールAから継承して新たな子ルールを生成した場合を表している。

3.6 突然変異

突然変異オペレーションとは、ある個体の染色体の一部をある確率に従い書き換えるものである。交叉と同様に、この突然変異の方法もいくつか種類があり、主なものとしては染色体の一部を同じ染色体の他の部分に移す転座法、染色体上である一定の長さの遺伝子を重複させる重複法、染色体の部分的な遺伝子配列の順序を入れ換える逆位法、染色体の一部を確率的に違う値に書き換える反転法などがある。

突然変異では、最適解探索のランダムな部分を与えており、特定の染色体パターンに収束することを防ぐことを目的としている。

本システムでは、それぞれの個体に、ある確率 P_{mut} で突然変異オペレーションを行ない、染色体の一部を違う値に書き換える。染色体は0, 1のバイナリー値をとるので、0なら1, 1なら0へ書き

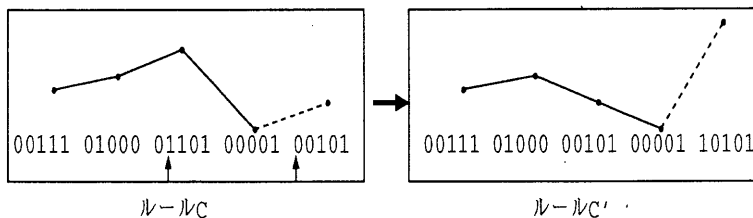


図5 突然変異による新たなルール生成の例

換えることになる。

図5には、突然変異オペレーションによってルールが書き換えられる例を示す。元のルールCの矢印が付されているビットパターン部分に書換えが起こりルールC'が生成される場合を表現している。

4 予測シミュレーション

適応的ルール生成システムによる株価時系列予測のシミュレーションを行なう。

ルール集団を構成するバイナリ値表現されたルールの個数は2000とし、10進数で表現された株価を2進数に変換する場合のビット数は30ビットとした。よって、1つのルールを表現する染色体の遺伝子記号列は全体で $(30 \times (p+1))$ ビットとなる。また、これらのルールは遺伝的アルゴリズムによって、その信頼度に従って新たなルール集団を構成することになる。

予測の対象となる株価時系列はさまざまな業種からランダムに選択した企業の株価、日経平均株価を用いた。

個別の企業株価は、東洋経済新報社のCD株価(1970年~1982年収録)の終値、日経平均株価は日経テレコンからダウンロードしたものを用いた。

図6には、日経平均株価に対する1日先の予測結果を示している。実線は平均株価の実現値、破線は適応的ルール生成システムによる予測値を表す。ルールの信頼度の評価、それによる新たなルールの生成には、株価データを図6の1~200サンプルまでの期間を用いており、201~300サンプルの期間は、予測を行なうためのルール評価には直接的には関与していないことになる。よって、1~200サン

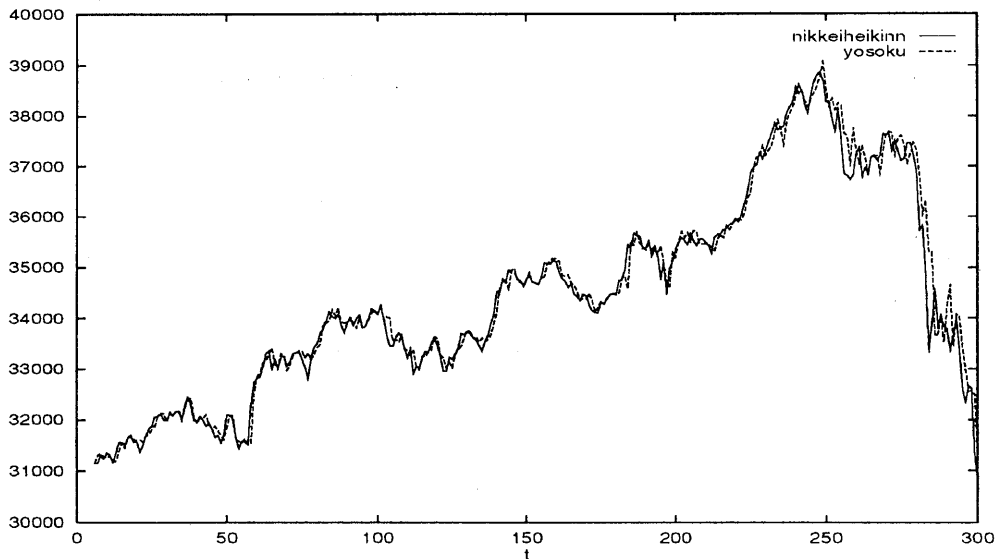


図6 本システムによる1日先の予測結果

ブル期間における予測は、本システムによる予測ルールが適切に生成されているか確認を行なう部分と考えることができ、201～300サンプル期間は生成されたルールがこの全体の株価パターンにどの程度適合しているかを検証することができる期間となる。

予測の誤差を全区間（1～300）における時系列の振幅（最大値－最小値）に対する予測値と実測値の差の絶対値の割合（％）として計算してみると、1～200期間では1.229％、201～300の期間では3.505％となり、ルール生成に直接用いた時系列パターンの予測の方が64％ほど予測誤差が小さいことが分かる。

同様な予測をいくつかの産業のランダムに選択した企業に適用した結果を表1にまとめている。日本化成、ユニチカは全期間において変動が大きかったため十分な適応が行なわれていないと考えられる。

全体的には、やはりルールの評価に用いた期間（1～200）の方が誤差は小さく、評価を行なわなかった期間（201～300）は予測誤差が比較的大きいようである。

比較のためARモデルによる1日先の株価の予測結果を表2にまとめている。ARモデルの次数は、パターンマッチシステムの条件部に用いる過去の株価数 p と同数とした。標本自己相関関数は1～200の範囲のサンプルを用いて計算し、その後、ユールウォーカー方程式によってパラメーターの推定をおこなった。

ARモデルの予測誤差に比べて本システムの予測誤差は、ルール評価に用いた期間（1～200、ARモデルの場合はパラメーター推定に用いた期間）の予測誤差の方がわずかに正確であるようだが、モデルの全体の時系列に対する適合性をみる期間（201～300）では、ARモデルの方が予測誤差は小さくなっている。理由としては、本システムはルール生成のための評価を行なう期間の株価の変動パターンにルール全体が特化してしまったためと考えられる。

	1～200期間の予測誤差	201～300期間の予測誤差
日経平均	1.229	3.505
マルハ	1.452	4.023
日本化成	3.668	4.767
ユニチカ	2.672	5.673

表1 本システムによる1日先の予測誤差

	1～200期間の予測誤差	201～300期間の予測誤差
日経平均	1.708	3.157
マルハ	2.272	3.422
日本化成	5.436	4.023
ユニチカ	2.108	2.293

表2 ARモデルによる予測誤差

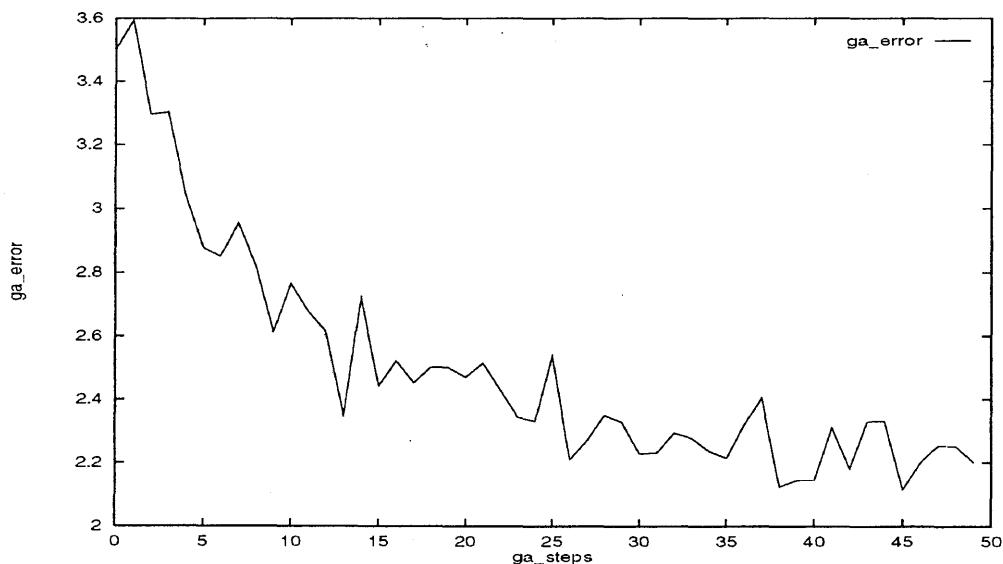


図7 GAエラーの変化

また、図7には、遺伝的アルゴリズムのサイクル数（ x 軸）についての予測値と実測値との平均2乗誤差（GAエラー）を示している。

15サイクルほどまでは、GAエラーは減少傾向にあるが、それ以降はわずかな減少傾向を示すものの、収束は遅くなっている。収束を速めるためには、ルール集合を表す個体数を増加することや、選択する割合を変更することが考えられるがどちらも、計算時間を大幅に増大させる可能性があるので注意が必要となる。

4.1 n 日先の予測シミュレーション

以上までは、1日先の予測を行ってきたが、次に、予測する日数を n 日先とした場合の予測シミュレーションを行う。

n 日先の予測方法としては、まず、 p 期前までの株価 ($S_{t-p}, S_{t-p+1}, \dots, S_t$) を用いて1日先の予測値 (\tilde{S}_{t+1}) を求め、その予測値を実際の株価 (S_{t+1}) の代わりとして、株価パターン ($S_{t-p+1}, S_{t-p+2}, \dots, S_t, \tilde{S}_{t+1}$) が外部環境から与えられたと仮定する。そして、与えられたと仮定した株価パターンと条件部がマッチしたルールが活性化され、そのルールの結論部を次の期（2日先）の予測値 (\tilde{S}_{t+2}) と考えることとする。同様な方法を繰り返すことで、 n 日先の予測を行なうことができる。

用いる株価時系列は日経平均株価とし、予測誤差の定義は1日先の予測と同じ、変動範囲に対する割合とした。表3には、適応的ルール生成システムとARモデルの1~200期間（適応的ルール生成システムではルールの評価を行なった期間、ARモデルの場合はパラメーター推定を行なった区間）の n 日先の予測誤差の比較結果を示している。

全体的に本システムがわずかに予測誤差が少なくなっているが、大きな差はなく、ARモデルと同程

	$n=1$	2	3	4	5	6	7	8
本手法	1.229	2.586	2.982	3.083	4.036	4.311	4.694	4.904
AR	1.708	2.477	3.102	3.698	4.140	4.522	4.807	5.087

表3 1～200期間の n 日先の予測誤差比較

	$n=1$	2	3	4	5	6	7	8
本手法	3.505	5.524	6.913	6.597	8.028	8.855	10.159	10.292
AR	3.108	4.881	6.056	7.611	8.900	10.119	11.196	12.427

表4 201～300期間の n 日先の予測誤差比較

度の予測傾向にあるといえる。同様に、4には、201～300の期間の予測誤差の比較を示している。本手法の方が比較的誤差は小さいが、7日先の予測で約10%の予測誤差に達している。

5 むすび

本論文では、人工知能の分野で、自動的プログラミングを行なうシステムとして考え出されたクラシファイアシステムの枠組を利用し、新しい株価時系列予測システムである適応的ルール生成システムを構築し、ARモデルとの予測誤差の比較を行なった。その結果、本システムは、新しい構想のもとの株価時系列予測を与えているが、株価のもつランダム性により、予測誤差はARモデルと同程度、あるいは、わずかに小さい値となることが示された。よって予測性能としては不十分であると考えられる。

今後の課題としては、本システムにカオス、フラクタル理論を取り入れたシステムを考察していきたい。

参 考 文 献

- (1) 北野宏明編：“遺伝的アルゴリズム”，産業図書，1993。
- (2) 北野宏明編：“遺伝的アルゴリズム2”，産業図書，1993。
- (3) Belew, R.: “Artificial Life : A Constructive Lower Bound for Artificial Intelligence”, IEEE Expert, Vol.6, No.1, Feb. 1991.
- (4) Fogel, L., Owens, A. and Walsh, M.: “Artificial intelligence through simulated evolution”, New York, John Wiley, 1966.
- (5) Goldberg, D.: “Computer-aided gas pipeline operation using genetic algorithms and rule learning”, Ph.D Thesis, University of Michigan, 1983.
- (6) Goldberg, D.: “Genetic Algorithms in Search, Optimization, and Machine Learning”, Addison-Wesley, 1989.
- (7) Langton, C.: “Artificial Life”, Addison Wesley, 1989.
- (8) Powell, D., Tong, S., Skolnick, M.: “EnGENEous: Domain independent, machine learning for design optimization”, Proc. of ICGA-89, 1989.
- (9) Robertson, G.: “Parallel Implementation of Genetic Algorithms in a Classifier System”, Davis, L. (Ed),

Genetic Algorithms and Simulated Annealing, Morgan Kaufmann Publishers, 1987.

- (10) Smith, S.F: "A Learning System Based on Genetic Adaptive Algorithms", Ph. D. Thesis, University of Pttsburgh, 1980.
- (11) Tanese, R.: "Distributed Genetic Algorithms", Proc. of ICGA-89, 1989.