

# 説得力のある時系列予測に関する研究

## A Research on Reliable Prediction of Time Series

伊藤友基<sup>†</sup>  
Yuki Itoh

長尾智晴<sup>†</sup>  
Tomoharu Nagao

### 1 はじめに

個人投資家の増加によって注目を浴びている株価などの経済時系列データや生活習慣病との関連性を発見できる可能性が高い人体に関する時系列データなど、我々の身の回りには様々な時系列データが存在し、それらの予測には大きな期待が寄せられている。これまでに神経回路網や進化論的手法による予測がされておりその有効性も示されている [1]。

本報告では、遺伝的プログラミング (GP) [2] を用いて、対象波形の近似関数を学習し、未知のデータに適用することで予測を行なった。また、学習期間の窓を動かしていくことで季節の違いなど時間の経過に影響されないシームレスなモデルを提案する。

さらに、従来は予測値の評価が MAPE や RMSE などによって単一的にされていたのに対し、本報告では、使用者が最も納得できる予測結果提示モデルを使用者に対するヒアリングを行なうことによって構築し、システムに組み込むことで、より説得力があり決定支援につながる時系列予測を目指した。

### 2 提案手法

#### 2.1 GPによる予測

例えば  $f(x) = x_1 \times x_2 + \sin(x_3)$  という関数は図1のように木構造で表すことが可能である。本報告では木構造を最適化する GP を用いて対象となる時系列データの近似関数を木構造によって構築して予測に用いた。木構造の1入力ノードには  $\{\sin, \cos, \ln, \log, \times 10, \div 10\}$  を、2入力ノードには  $\{+, -, \times\}$  を使った。ただし、入力値は 0.1 ~ 0.9 の範囲に正規化した値を用いた。

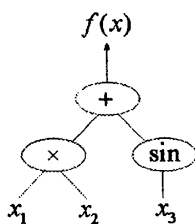


図1 木構造で表される数式

#### 2.2 シームレスなモデルの提案

図2のように学習・予測という1試行を学習期間の窓を動かしながら時間の経過と共に逐次繰り返す。ここで、新たな予測試行に移る際、1step 動かしただけでも

かわらず、前試行で学習した結果を破棄し、全個体を初期化して初めから学習を行なうのは、効率も悪く、学習したトレンドを維持する点においても不適切であると考えられる。

そこで、予測実績が著しく悪い場合には、個体を全て初期化するが、一定以上の実績がある場合、一定数の個体を保存しておき、次の試行へと引き継ぐことで、学習したトレンドを保持しつつ、時間の経過に従って新たなトレンドへ自動的に対応していくシームレスなモデルを提案する。例えば、季節に応じた予測器を構築した場合、従来では季節の変わり目にはどの予測器を用いるべきか非常に曖昧であったが、シームレスなモデルにすることで、時間の変化による影響に対して自動的に対応が可能であると考えられる。

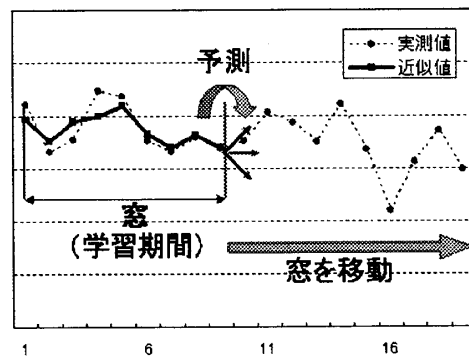


図2 シームレスな処理

#### 2.3 説得力のある予測

ここで、説得力があるとは、“予測器の使用者がより納得して予測値を採用できる結果提示が可能であること”とする。最適な結果提示方法は使用者によって異なるため、各使用者に応じて検討する必要がある。そこで、本報告ではある時系列データの予測者であるエキスパートを使用者と想定してヒアリングを行ない有効な予測提示方法を検討した。以下に、今回対象とした使用者に対して有効な特徴を示す。

- 様々な方向性から求まる複数の予測値の提示
- 予測値に幅をもって提示
- 結果が近い過去の事例を示す

以上の調査結果から、独立して学習する GP の個体プールを複数個用意し、学習期間の幅や重み率、入力データなどの設定をプールごとに変えて個性を出すことで、全体としては様々な方向性から求められた複数の予測値が得られるモデルとすることにする。

<sup>†</sup> 横浜国立大学 大学院環境情報学府, Graduate School of Environment and Information Sciences, Yokohama National University

### 3 実験

これまでに述べた条件において1試行を、

- 個体数 100 個体
- 学習世代数 5000 世代
- 世代交代方法 MGG [3]
- 学習期間 20 日
- 重み率 0.955

として学習・予測を行なった。ただし、学習期間や重み率は経験的に決めた。適応度は式(1)のように実測値  $f(t)$  と近似値  $f^*(t)$  の誤差に、予測日  $W$  からの距離に応じた重み率  $w(\leq 1)$  によって重み付けした値とした。

$$fitness = \frac{1}{W} \sum_{t=1}^W w^{(W-t)} |f(t) - f^*(t)| \quad (1)$$

本実験では入力変数が異なる4つの予測器を定義し、予測器ごとに5回実行した平均値で精度を評価する。ここでは、次の試行に移る際に引き継ぐエリート個体以外の選択方法を、進化論的手法におけるトーナメント選択、上位個体から順番に選択、ランダム選択、プール内から平均的に選択する各場合および毎回初期化する場合について実験を行なう。得られた結果を比較し、試行移動時の個体の引き継ぎの妥当性および適切な個体選択方法も検討する。さらに、最良の結果が得られた選択方法を用いた実験結果において予測器ごとの精度の検討を行なう。

### 4 実験結果とその考察

#### 4.1 個体の引き継ぎ

次の試行に移る際、一定数の個体を次の試行へ引き継ぎながら学習を行なうことで、トレンドの連続性を保ちながら、移り変わるトレンドを自動で学習できると考えられる。本報告では、誤差実績が3%未満の場合に、全個体の3分の1を引き継いだ。図1に、ある時系列の2003年度のデータを、5種類の個体選択方法で予測した結果を示す。

表1 各個体選択方法による予測誤差 (%)

	通年	春・秋	夏	冬
引き継ぎ無し	2.29	2.84	2.04	2.01
上位個体から選択	1.90	2.52	1.55	1.64
トーナメント選択	1.92	2.61	1.59	1.52
ランダム選択	1.88	2.48	1.58	1.57
平均選択	1.86	2.33	1.64	1.63

個体を引き継ぐことによる精度の向上は明らかであり、本手法は有効であると言える。また、GPにおいてエリート個体だけでなく、適応度の低い個体も含めてプール全体が進化の結果と言えるので、全個体から満遍なく選択した場合の精度が最も良好であると考えられる。

#### 4.2 予測誤差

上記の平均選択の場合の各予測器と平均値の予測誤差を表2に示す。

対象データは、予測誤差は人間では3%、神経回路網による既存研究では2%程度であり、本手法では2%以下

表2 各予測器ごとの予測誤差 (%)

	通年	春・秋	夏	冬
予測器1	1.95	2.48	1.73	1.66
予測器2	1.97	2.51	1.73	1.69
予測器3	1.90	2.26	1.66	1.82
予測器4	2.05	2.68	1.78	1.72
平均値	1.86	2.33	1.64	1.63

の結果を得られたことから、予測器として十分な性能があると言える。

#### 4.3 結果の提示

最後に本報告で検討した結果提示方法の概要を図3に示す。個体を初期化するランダム値を変えて5回行った全ての値をプロットすることで、予測値の範囲として示すことが出来た。回帰分析のように予測値を1つだけ示す場合と比較して、エキスパートの方からも高い評価を得られた。

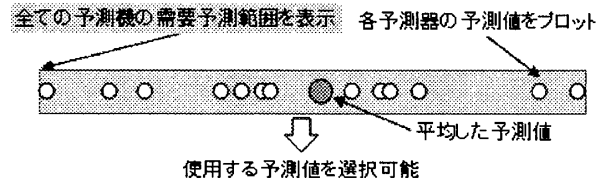


図3 結果提示方法

### 5 まとめと今後の課題

本報告では、一定以上の精度の出るシームレスな予測器を構築し、どのような結果提示方法が、最も説得力があり、使用者が納得して使用することが可能であるか検討した。ただし、結果提示対象を特定の狭い対象に絞って検討しているという問題点があり、今後は、対象とするドメインを広げていく必要があると言える。その中で、後者の説得力に関して、使用者の置かれている背景や要望を、オントロジー [4] などの手法によって明確にモデル化し、使用者が最も満足できるような結果の提示方法を検討することで、実際に使用者の決定支援に用いることができるような予測器を構築していきたい。

#### 参考文献

- [1] Baba,N. and Kozaki,M.: "An intelligent forecasting system of stock price using neural network", Proceedings of IJCNN, 1, pp.371-377 (1992)
- [2] J.R.Koza: "Genetic Programming on the Programming of Computers by Means of Natural Selection", MIT Press (1992)
- [3] 小林 重信 他: "遺伝的アルゴリズムにおける世代交代モデルの提案と評価", 人工知能学会誌, Vol.12, No.5, pp.1-10 (1997)
- [4] 溝口 理一郎 他: "オントロジー工学序説-内容指向研究の基盤技術と理論の確立を目指して", 人工知能学会誌, Vol.12, No.4, pp.559-569 (1997)