

## 進化的計算手法を用いた多変量予測モデルの動的最適化

佐藤 春樹<sup>†</sup> 鈴木 智也<sup>††</sup><sup>†</sup>同志社大学工学部 <sup>††</sup>同志社大学理工学部

## 1 はじめに

近年、コンピュータの普及により、プログラムによって様々なデータを自動的に予測する試みがなされている。株価予測を例に挙げれば、自動で将来株価を予測し、さらに自動で株式売買をするシステムを構築するコンテストも行われている [1]。株価や為替価格のような経済システムにおいては、企業毎の株価や為替交換ペアに応じて複数の銘柄が存在し、それぞれが互いに影響し合うことで複雑な挙動を示している。このような多変量システムの挙動をうまく予測するには、予測モデルを適切に設定する必要がある。

本研究では、予測モデルの最適化として、予測に使用する銘柄と使用しない銘柄を選別する手法を提案する。その手法として、進化的計算手法である遺伝的アルゴリズム (以後 GA) [2] を用いて、過去のデータの予測精度を最良にする予測モデルを探索する。この探索された予測モデルを用いて、新規データに対する予測を行う。

本手法の有用性を調べるために、数値シミュレーションを行う。予測対象データとして、複雑系を模擬する数理モデルや実際の外国為替取引価格を用いる。さらに、実際のシステムを予測対象とする場合、外部からの影響によりシステムの構造が動的に変化している可能性を考慮する必要がある。そこで、毎回の予測において予測モデルを最適化し直す方法も検討する。このような動的最適化による効果についても数値シミュレーションを通じて検証する。

## 2 GA を用いた多変量システムの非線形予測

多変量システムにおいて観測できる要素数を  $N$ 、各要素の時系列データを  $x_i(t)$  ( $i = 1, 2, \dots, N$ ) とすると、システム全体の振舞いは以下のように記述される。

$$V(t) = \{x_1(t), x_2(t), \dots, x_i(t), \dots, x_N(t)\} \quad (1)$$

本研究では、 $V(t)$  の将来変動を予測するために、近傍の振舞いを参考にすることで、局所線形近似的に非線形予測を行う [3]。そのため  $V(t)$  の近傍点  $V(k_n)$  を  $t-L \leq k_n < t$  より探し、その 1 ステップ後の振舞いを平均化することで、 $V(t+1)$  の予測値  $\tilde{V}(t+1)$  を得る。

$$\tilde{V}(t+1) = \langle V(k_n + 1) \rangle_n \quad (2)$$

しかし、一般的に式 (1) を適切に構成することは容易ではない。観測した変数が、システムに内在する変数であるのかは不明であるし、もし内在する変数だとしても予測対象とする変数と直接関連しているとは限らない。

つまり、何らかの方法で全変数の中から予測に必要な変数を厳選する必要がある。例えば、変数  $i$  が変数  $i_1, i_2, i_3$  からの影響を受けていれば、変数間の因果構造は次式のように書ける。

$$x_i(t+1) = \mathbf{F}(x_i(t), x_{i_1}(t), x_{i_2}(t), x_{i_3}(t)). \quad (3)$$

ここで  $\mathbf{F}$  は関数、 $i$  は目的変数、 $i_{1\sim 3}$  は説明変数である。この場合は、式 (1) を

$$V(t) = \{x_i(t), x_{i_1}(t), x_{i_2}(t), x_{i_3}(t)\} \quad (4)$$

に修正し、式 (2) を適用する方が優れた予測精度を期待できる。

予測対象変数  $i$  と直接的に相互作用する要素集合  $\{i_j\}$  は、 $2^N$  通りの可能性がある。この組合せ総数は、要素数  $N$  に応じて組合せ爆発を起こすので、式 (3) のような因果構造を推定し、式 (4) のように予測モデルを最適化する問題は組合せ最適化問題となる。そこで我々は、先攻研究 [4] において、GA を用いることで、この問題の解法を検討してきた。

まず GA で用いる遺伝子型を  $g_i = \{11001 \dots\}$  のように 2 値表現し、 $g_i(j) = 0$  であれば変数  $i$  の予測の際に、式 (1) の  $V(t)$  として変数  $j$  を用いないことを意味し、 $g_i(j) = 1$  であれば変数  $j$  を用いることを意味している。さらに、各遺伝子型の適合度を求めるために、それぞれの  $V(t)$  を用いて、過去に観測されたデータを予測することで予測精度を算出し、その予測精度を適合度とする。交叉や突然変異など GA のアルゴリズムによって世代を進めると遺伝子型が最適化され、これを基にシステムの因果構造や最適な予測モデルを推定することができる [4]。

## 3 提案手法の有用性の検証

前章の手法の有用性を検討するために、実際の多変量システムとして、1996 年の外国為替市場における 25 種の取引価格を予測した。各取引価格は 30 分毎に記録され、時系列データを構成している。ただし実システムは一般的に予測が難しいことから、1 ステップ後の変動が上昇するか、または下降するかを予測した。つまり  $\hat{x}_i(t+1) - x_i(t) > 0$  であれば上昇、 $\hat{x}_i(t+1) - x_i(t) < 0$  であれば下降と予測し、その的中率を予測精度 [%] とした。

さらに、次式の数理モデル [5] に対しても予測を行った。

$$x_i(t+1) = \mathbf{F} \left( (1 - \epsilon)x_i(t) + \frac{\epsilon}{N_i} \sum_{j \in \{i_j(p)\}} x_j(t) \right) \quad (5)$$

Dynamical Optimization for Multivariate Nonlinear Prediction with Evolutionary Computation

<sup>†</sup> Haruki Sato (bte7117@mail4.doshisha.ac.jp)

<sup>††</sup> Tomoya Suzuki (tsuzuki@mail.doshisha.ac.jp)

Department of Information System Design, Faculty of Engineering, Doshisha University (<sup>†</sup>)

Department of Information System Design, Faculty of Science and Engineering, Doshisha University (<sup>††</sup>)

1-3 Tatara-Miyakodani, Kyotanabe-shi, Kyoto 610-0321, Japan

$N_i$  は要素  $i$  と結合している要素数,  $\epsilon$  はその結合強度を表す. ただし各要素の結合パターンはレギュラグラフだけでなく, WS モデル [6] を導入することで, スモールワールドネットワークやランダムネットワークに変更可能にした. WS モデルにおけるノードのランダム再結合変数を  $p$  とすると,  $\{i_j(p)\}$  は要素  $i$  と結合する変数  $j$  の集合を表している. また関数  $F$  として, 広く研究されているロジスティック写像  $F(x) = 1 - ax^2$  を採用した.

実際のシステムを対象とする場合, 外部の影響によってシステムの構造が変化する可能性がある. もし構造変化が起これば, 予測モデルの当てはまり悪くなるので予測精度は低下する. その対策として, 時間の経過に伴い新規データが観測される度に, 予測モデルを最適化し直すのが理想的であるが, 計算コスト観点から毎回の最適化は困難なケースも考えられる. そこで本研究では, 過去 20 回の予測精度が閾値  $\theta$  [%] を下回ったら, 再び GA にて因果構造を推定し直し, 予測モデルを動的に最適化する実験を行った. ただし, 図中の C1 は, 最初の 1 回のみ GA を適用した後, 予測モデルを固定する方法を意味し, C2 は毎回の予測において GA にて予測モデルを最適化し直す方法を意味している. さらに, GA による動的最適化において計算コストを低減させるために, 前回の最適化において適合度の良い遺伝子型の 30 [%] を, 今回の初期遺伝子型として再利用し, GA の世代数を半分にする方法も実践した.

実データおよび数理モデルに対する予測精度を図 1 に示す. GA を用いず観測された全変数を用いて予測した場合 (つまり式 (1) を適用) と GA を用いて予測モデルを最適化する手法 (提案法) の比較においては, 提案法の有用性は明らかである. さらに, 為替取引価格の予測では, 閾値  $\theta$  [%] を大きくして動的最適化の回数を増やすほど予測精度が向上するので, 動的最適化の有用性も確認できる. さらにこの結果は, 外国為替市場の構造が動的に変化している事実を裏付けている. このように実システムを予測対象とする場合には, 本研究で示す動的最適化法が有効である.

しかし, 数理モデルの予測においては, 閾値  $\theta$  [%] を大きくしてもあまり予測精度は向上しない. これは, 数理モデルの構造が変化しないことを表しており, 式 (5) より当然の結果である. このような手法を通じて, 解析対象のシステムが構造変化しているか否かを分析できることも興味深い.

適合度の良い遺伝子型を再利用する試みに関しては, 数理モデルの予測において僅かな予測精度の低下が確認できる. しかし, GA の世代数を半減した上での予測精度であるので, 計算コストを抑えたい場合には, 有用な手法だと考えられる.

#### 4 まとめ

本研究では, 観測によって得られた多変量データを厳選することにより, 予測モデルを最適化する手法を検討した. その最適化手法として, GA が応用できることを数値シミュレーションにより示した. また, 外国為替市場などの経済システムのように外部からの影響を受けやすいシステムは, 構造が動的に変化する可

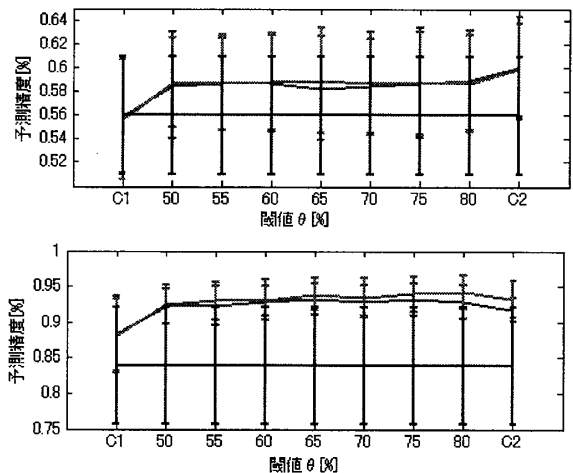


図 1: GA にて予測に用いる変数を厳選した場合 (赤線と緑線) および全変数を用いて予測した場合 (青線) の比較. 上図は, 為替取引価格 (25 銘柄) を予測した場合, 下図は, 数理モデル (式 (5)) を予測した場合である. ただし, 赤線は動的最適化において, 前回の適合度の良い遺伝子型の 30 [%] を初期遺伝子型として再利用する方法であり, 緑線は再利用をせず初期遺伝子型を全て変更し, GA を最初からやり直す方法である. なお, 各実線は各銘柄の予測精度の平均値を示し, 各エラーバーは予測精度の標準偏差を示している.

能性がある. そのようなシステムに対しては, 予測を行う度に予測モデルを最適化し直すことで, 予測精度を向上できる事を示した.

今後の課題として, より実践的な観点から本提案手法の有用性を検証すべく, 投資シミュレーションを行う必要がある. つまり, 投資による利益率の観点から, 本提案手法の有用性を評価する予定である.

なお本研究の一部は, 日本学術振興会科学補助金若手研究 (B) (No.20700217) およびケン・ミレニアム株式会社物理学者支援プログラムの援助により行われました.

#### 参考文献

- [1] 自動株式売買ロボット・カブロボ: [http:// www.kaburobo.jp/](http://www.kaburobo.jp/), (参照日 2008/1/8).
- [2] 伊庭齊志: “遺伝的アルゴリズムの基礎,” オーム社, 1994.
- [3] E. N. Lorenz: “Atmospheric predictability as revealed by naturally occurring analogues,” J. Atmos. Sci., Vol.26, pp.636–646, 1969.
- [4] T. Suzuki, Y. Ueoka, H. Sato: “Combinatorial Optimization for Multivariate Nonlinear Prediction,” Proc. of NCSP’09, 2009.
- [5] Y. Ueoka, T. Suzuki, T. Ikeguchi, Y. Horio: “Efficiency of Statistical Measures to Estimate Network Structure of Chaos Coupled Systems,” Proc. NOLTA’08, 2008.
- [6] D. Watts and S. Strogatz: “Collective dynamics of ‘small-world’ networks.” Nature, vol.393, pp.440–442, 1998.