

自己適応型差分進化法に基づく AR-HMM

小野 景子^{1,a)} 鳥山 直樹¹ 古川 雄大¹ 折登 由紀子^{2,b)}

概要: 時系列データにおける潜在的な状態推定は、自己回帰隠れマルコフモデルを用いることが一般的である。この推定問題は2段階の最適化問題であり、回帰モデルにおいて参照する過去のデータ数(次元)と各時点の観測データに対する重み係数を同時に最適化する必要がある。しかし、自己回帰隠れマルコフモデルは設計変数決定に多くの計算量が必要であり、全ての候補解の比較は非効率である。そこで、我々は自己回帰隠れマルコフモデルにおける2段階の最適化を同時に行えるフレームワークを提案し、その最適化手法として、自己適応型差分進化法を提案する。提案差分進化法では、異なる次元の持つ個体が一つの母集団を形成し、進化により最適な次元と重み係数を持つ個体が生き残り、適応的な生存戦略を実現する。数値実験では、状態が複数ある人工データの推定、株価のリアライズボラティリティーの推定および訪日外客数の推定を対象とし、提案手法が従来の回帰モデルより状態推定性能が高いことを示す。

キーワード: 自己回帰隠れマルコフモデル, 2段階最適化問題, 差分進化法, 適応型探索

AR-HMM based on Self-Adaptive Differential Evolution

KEIKO ONO^{1,a)} NAOKI TORIYAMA¹ YUTA FURUKAWA¹ YUKIKO ORITO^{2,b)}

Abstract: The estimation of a hidden state based on time series data for a given period is a problem associated with the development of an auto-regression hidden Markov model. This is a two-step optimization problem where the order of past data (dimensions) used in the regression model is decided, and weighted coefficients for observed data at each point in time are determined. However, the auto regressing hidden Markov model used in this study for regression requires large computational effort to determine design variables due to the use of the hidden Markov. For such an optimization problem, we propose a self-adaptive differential evolution with a framework capable of simultaneous optimization of the first and second steps. The proposed method takes an approach where in individuals that represent the solution space of different orders in a population are generated. Further, it retains individuals with an order that has a large number of good solutions at a high probability through evolving. Numerical tests will involve performance validation using estimated artificial data of several states, realized volatility of the stock, and artificial as well as actual data of inbound visitors to Japan, and they will demonstrate that the optimization of the regression model by the proposed method is more effective than that by a conventional method.

Keywords: AR-HMM, Two-step optimization problem, Differential Evolution, Adaptive search

1. Introduction

ある期間の傾向やイベント発生による変動を分析など、

¹ 龍谷大学
Department of Electronics and Informatics, Ryukoku University, Japan

² 広島大学
Graduate School of Economics, Hiroshima University, Japan

^{a)} kono@rins.ryukoku.ac.jp

^{b)} orito@hiroshima-u.ac.jp

近年、様々な場面で時系列データ解析の必要性が高まっている。時系列データ解析は、主に経済時系列データの分析と予測のために発展し、一般的に、過去のデータ時系列からその延長線上の将来のデータを推定する回帰モデルを構築することで、過去のデータの分析と将来のデータの推定を行う。このような回帰モデルとして、自己回帰(AR)モデル、移動平均モデル、自己回帰移動平均モデル、自己回帰和分移動平均モデル、分散自己回帰モデル、一般化分

散自己回帰モデルなど様々なモデルが提案されている [1] . また近年、時系列を異なる状態毎に捉えるスイッチングモデルが提案されている [2] . このスイッチングモデルの一つとして、AR モデルに隠れマルコフモデルを適用した AR-HMM が提案されており、本研究ではスイッチングモデルのひとつである AR 隠れマルコフモデル (AR-HMM) に着目する . この AR-HMM は、経済時系列だけでなく、音声時系列や気象時系列などの時系列データの状態推定に適用されている [3] [4] . 一般に、回帰モデルは 1 時点先のデータを推定することを目標とし、過去データの次数 (推定に用いる過去の時点数) を定め、次に各時点のデータに対する重み係数をデータから最尤推定で学習することで構築される . つまり、回帰モデルの最適化問題は、まず解空間の次元数を決定し、その下で設計変数を決定する二段階の最適化問題とみなすことができる . 一方、モデルの評価はモデルの構築後に行うため、一段階目と二段階目の最適化を同時に行うことが難しい最適化問題であり、通常、全ての解を比較し、モデルの評価関数値が最小となるの回帰モデルが選択される . 本研究では、これらを同時最適化する問題を多段階最適化問題と呼ぶ .

多段階最適化問題は、設計者が予め経験に基づいて定めるハイパーパラメータと設計変数の両方を最適化する問題と考えることが可能であり、広く実問題に存在する問題領域であり、その解法の開発は重要である . 本研究で対象とする AR-HMM のような時系列データに対する回帰モデルの推定は、状態数と次数が増えた場合、最適な値の組み合わせは指数的に増加するため、効率的な探索は難しい .

一方、進化計算は多点探索と基本とし、定めた目的関数を最小にする最適な探索点を自然淘汰により求める手法である . 一般に、設計変数の次元は固定であるが、設計変数の次元および設計変数値の両方を同時進化させることで、多段階最適化問題の解法に有効である可能性がある . そこで、本研究では、多段階最適化問題に対する手法開発の第一歩として、AR-HMM のための二段階最適化問題の解法として、進化計算を用いたフレームワークを提案する . 自然淘汰で評価が良い探索点が生き残り、その探索点の周辺領域で更なる性能向上のための探査が行われるため、重点探索により計算コストが減る . そのため、進化計算を用いたフレームワークは AR-HMM の状態数と次数が増えた場合においても有効に働くと考えられる . また、状態数と次数が独立でないため、高い局所探索能力を持つ最適化手法が適する . そこで、本研究では、高次元の最適化問題の解法として有効な差分進化法 (Differential Evolution:DE) を本手法の最適化手法として用いる . 提案する自己適応型 DE の個体は状態数 × 次数の設計変数を持ち、最適な次数とその重みは提案する生存戦略に基づき進化の過程で獲得する .

本論文の構成は次の通りである . 2 章では、AR モデル、

及び AR-HMM について述べる . 3 章では、AR-HMM の問題点を議論し、状態数と次数を同時最適化するための自己適応型 DE を用いたフレームワーク (SaDE-AR-HMM) を提案する . 4 章では、人工的に作成したテストデータと 2 つの実データを用いて提案法の性能を評価し、分析結果について述べる . 5 章はまとめを述べる .

2. 時系列データの分析モデル

時系列データの振る舞いを捉える最も一般的な方法は、過去の観測値と現在の観測値の相関を推測することであり、自己相関を持つ時系列データの推定モデルの代表として AR モデル [5] が提案されている . さらに、時系列データの潜在する状態を捉えるモデルとして、AR モデルに隠れマルコフモデルを適用した AR-HMM [2-4] が提案されている . 本研究では DE を用いた AR-HMM の性能向上を目指す . 本章では、まず、AR モデル、および AR-HMM のアルゴリズムについて述べる .

2.1 Auto Regressive(AR) Model

時系列データ $\mathcal{X} = \{x_1, \dots, x_N\}$ に基づき、現在と過去の観測値の関係を推定することを考える . 自己相関を持つ時系列データの振る舞いを捉えるためのモデルとして、自己回帰 (AR) モデルが提案されている . AR モデルは、同一時系列データの現在の観測値から M 期前までの観測値の線形和によって将来の予測を行うモデルである . 時点 n の将来データ x_n を推定する次数 M の AR(M) モデルは次式で表される .

$$x_n = \sum_{m=1}^M a_m x_{n-m} + v_n, \quad v_n \sim N(0, \sigma^2) \quad (1)$$

ここで a_m は自己回帰係数、 v_n は平均 0、分散 σ^2 の正規分布から生成される乱数である . $\phi = \{a_1, \dots, a_m, \dots, a_M\}$ とした場合、AR モデルの尤度関数は次式で与えられる .

$$p(\phi, \sigma^2) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} \left(x_n - \sum_{m=1}^M a_m x_{n-m}\right)^2\right] \quad (2)$$

AR モデルのパラメータのはユールウォーカー法 [6] を用いることで解析的に解くことが出来る . (2) 式の対数を a_m で微分すると次式を得る .

$$-\frac{1}{\sigma^2} \sum_{n=1}^N \left(x_n - \sum_{l=1}^M a_l x_{n-l}\right) x_{n-m} = 0 \quad (3)$$

$$\sum_{n=1}^N \sum_{l=1}^M a_l x_{n-l} x_{n-m} = \sum_{n=1}^N x_n x_{n-m}$$

ここで $\sum_{n=1}^N x_{n-m} x_{n-l}$ を自己共分散 $C_{|l-m|}$ とする . 各パラメータ a_1, \dots, a_M についても同様に微分すると、次式になる .

$$\begin{pmatrix} C_0 & C_1 & \dots & C_{M-1} \\ C_1 & C_0 & \dots & C_{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{M-1} & C_{M-2} & \dots & C_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_M \end{pmatrix}$$

上記の連立方程式を解くことによって、 ϕ を解析的に求めることが出来る．また分散 σ^2 は次式で与えられる．

$$\sigma^2 = \frac{1}{N+M} \sum_{n=1}^N (x_n - \sum_{m=1}^M a_m x_{n-m})^2 \quad (4)$$

次に、AR モデルの次数決定法について述べる．AR モデルをはじめとする回帰モデルでは、次数がモデルの推定性能に大きな影響を与える．次数が小さすぎた場合、モデルの表現能力が低くなり適切な推定が行えなくなる．一方、次数を大きくするとモデルはより複雑な振る舞いを表せるようになるが、次数を過剰に大きくすることはオーバーフィッティングの原因になる．また、推定すべきパラメータが増加するためパラメータの推定が難しくなる．そこで本研究では、これらのバランスをとるために、モデルの評価関数として AIC(赤池情報量基準) [7] を用いる．AIC は以下の式で与えられる．

$$AIC = -2 \sum_{n=1}^N \ln p(x_n | \phi, \sigma^2) + 2m. \quad (5)$$

ここで、 m は AR モデルの次数である．一般的には、次数を $m = 1, 2, 3, \dots$ のように変化させながら AIC が最小となるような M を求める．

2.2 AR-HMM

時系列データはその振る舞いが突然大きく変化し、それまでとは異なる自己相関を持つ変動をする場合がある．そのような場合、AR モデルを用いて時系列データを分析するためには、変動の異なる期間を捉え、それぞれの期間ごとに異なる AR モデルを適用する必要がある．しかしながら、手作業で、変動の異なる期間を捉えることは難しいため、複数の状態を持った時系列データの変動を捉えるモデルとして、AR モデルに隠れマルコフモデルを適用した AR-HMM が提案されている [2]．AR-HMM のアルゴリズムを以下に示す．

時点 $n = 1, 2, \dots, N$ における時系列データを $\mathcal{X} = \{x_1, \dots, x_N\}$ 、状態変数を $\mathcal{Z} = \{z_1, \dots, z_N, \dots, z_N\}$ とする．ここで状態変数 z_n は一対 K 符号化法により第 k 要素のみが 1 であり、その他の要素が 0 の K 項ベクトルとして表現される．図 1 に、状態数 $K = 3$ における AR-HMM のイメージを示す．図 1(a) は状態変数 \mathcal{Z} の遷移を表しており、初期状態 z_{11} から $z_{21} \rightarrow z_{32}, \dots, z_{n3}$ へと遷移している．図 1(b) は、推定された状態変数 z_{nk} から観測データ x_n が生成されるイメージを表している．

観測データと状態変数の同時分布は以下の式で表される．

$$p(\mathcal{X}, \mathcal{Z} | \theta) = p(z_1 | \pi) \left[\prod_{n=2}^N p(z_n | z_{n-1}, \mathbf{A}) \right] \prod_{n=1}^N p(x_n | z_n, \Phi) \quad (6)$$

ここで、 $\theta = \{\pi, \mathbf{A}, \Phi\}$ はパラメータの集合であり、 $\pi = \{\pi_1, \pi_2, \dots, \pi_K\}$ は初期遷移確率、 \mathbf{A} は $K * K$ の遷移確率行列であり、 A_{jh} は時点 $n-1$ から n において状態が j から h へ遷移する確率である． $\Phi = \{\phi_1, \phi_2, \dots, \phi_k, \dots, \phi_K\}$ は各状態の AR モデルの自己回帰係数であり、 $\phi_k = \{a_1, a_2, \dots, a_{M_k}\}$ である． M_k は状態 k における AR モデルの次数である．次に、(1) 式の尤度を最大化するために EM アルゴリズムを導入する．状態変数 z_n の周辺事後分布を $\gamma(z_n)$ 、2 つの連続した状態変数の同時事後分布を $\xi(z_{n-1}, z_n)$ とし、更新前のパラメータ集合を θ^{old} とすると、パラメータ集合 θ を更新する関数 $Q(\theta, \theta^{old})$ は次式となる．

$$Q(\theta, \theta^{old}) = \sum_{k=1}^K \gamma(z_{1k}) \ln \pi_k + \sum_{n=2}^N \sum_{j=1}^K \sum_{k=1}^K \xi(z_{n-1,j}, z_{nk}) \ln A_{jk} + \sum_{n=1}^N \sum_{k=1}^K \gamma(z_{nk}) \ln p(x_n | \phi_k) \quad (7)$$

E ステップでは θ^{old} を用いて、 $\gamma(\cdot), \xi(\cdot)$ を計算し、M ステップでは θ を最大化する．

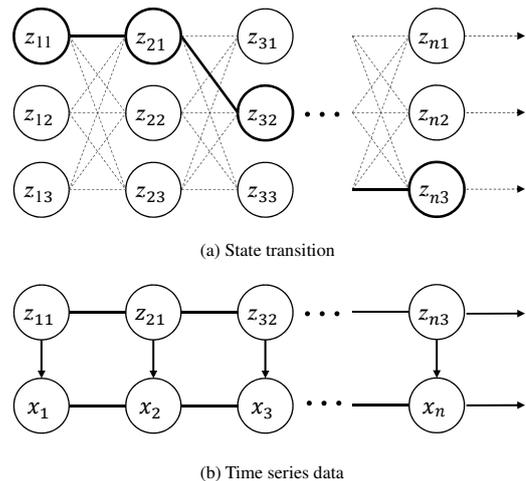


図 1 $K = 3$ の AR-HMM のグラフィカルモデル

AR-HMM では K 個ある AR モデルの次数の推定を行う必要があるため、AIC は以下の式で計算する [7]．

$$AIC = -2 \sum_{n=1}^N \sum_{k=1}^K \gamma(z_{nk}) \ln \{p(x_n | \phi_k, \sigma_k^2)\} + 2m_k. \quad (8)$$

ここで m_k は状態 k における AR モデルの次数である．

3. SaDE-AR-HMM

AR-HMM は時系列データには潜在的な状態が関係している, 例えば, データの変動が大きい, 小さいという変化を状態変数 z_n を用いて捉え, データの推定に用いる. 多くの時系列データはそのような複数の状態を含むため, AR-HMM は非常に有効な手法であることが知られている. しかしながら, 次数, 自己相関係数, 状態数などモデルの性能に影響するパラメータが多くある.

本章では, AR-HMM における問題を示し, またそれらの問題を解決するために AR-HMM にを適用した SaDE-AR-HMM を提案する. まず, AR-HMM における 2 つの問題を示す.

問題 1: AR モデルのパラメータの推定

AR-HMM では, 状態数 K における AR モデルのパラメータ $\Phi = \{\phi_1, \phi_2, \dots, \phi_K\}$ を最適化する必要がある. ここで $\phi_k = \{a_1, \dots, a_{m_k}\}$ である. すなわち, 推定すべきモデルパラメータの数は状態数及び AR モデルの次数に比例するため, 状態数や次数が増加した場合, 一般的な解法では良質な解が得られない場合がある.

問題 2: AR モデルの次数の決定

AR-HMM では, 各状態の AR モデルにおける適切な次数 $m_1^*, m_2^*, \dots, m_K^*$ を決定する必要がある. 対象となる時系列データに対し, 十分な専門知識がある場合には, それらをモデルの事前知識として用い, 次数の決定が可能である. しかしながら, 事前知識がない場合には AR-HMM モデルの状態 k における最適な次数と, 次数に応じたモデルパラメータを AIC の評価に基づいて学習する必要がある.

問題 1 に対しては, サンプリング法などを用いることで良質な解を推定することが出来る. また, 問題 2 に対しては, 状態数 K と AR モデルの次数の最大値 M について, 全ての組み合わせについてパラメータを推定し, その中で AIC が最も低くなるようなモデルを選択可能である. しかしながら, 次数や状態数が増加した場合, 問題 1 と 2 の解決には, 非常に多くの計算コストが必要である.

一方, 自然淘汰の仕組みを上手く用い, 探索が進むにつれて評価値が良い領域を集中的に探索する手法に進化計算手法がある. このように, 進化計算手法は全ての解候補を評価する必要がないため, 上述の問題 1 と 2 を解決する, AR-HMM の高性能化に適するといえる. また, 次数, 状態数および自己相関係数は依存関係が存在する. そのため, 本研究では, 進化計算手法の中でも, 高い探索性能をもつ DE を用いて, それらの最適化を試みる. 一般に, DE は複数の選択個体の差分ベクトルをもとに探索方向を決定するため, 個体の次元数は同じである. しかしながら, 状態

数 $k = \{1, 2, \dots, K\}$, 次数 $m = \{1, 2, \dots, M\}$ と自己相関係数を同時に最適化するためには, 異なる次元数の設計変数をもつ個体を同時に扱う必要がある. そこで, 異なる次元数の設計変数を有し, それらを学習可能な自己適応型差分進化法に基づく AR-HMM (SaDE-AR-HMM) を提案する.

提案法では, AR モデルのパラメータ Φ を個体の設計変数として定義し, AIC が最も低い最良個体をパラメータの更新に用いる. ここで, 各個体が持つ設計変数の数は AR モデルの次数を表す. 母集団 $\mathcal{I} = \{I_1, \dots, I_i, \dots, I_I\}$ において, 状態数が k の個体 I_i は設計変数 $\{\phi_1, \dots, \phi_k\}$ をもち, $\phi_k = \{a_1, \dots, a_{m_k}\}$ である^{*1}. 各状態と次数の組み合わせの総数は $K \times M$ であり, C_1 を $(k, m) = (1, 1)$ の, C_{KM} を $(k, m) = (K, M)$ のである個体サブ集合とした場合, 母集団 $\mathcal{I} = C_1 \cup \dots \cup C_\ell \cup \dots \cup C_{KM}$ である. このとき, E 個体のエリート集団 \mathcal{E} に基づき個体生成分布 P を

$$P(k, m|\mathcal{E}) = \sum_{\ell} \frac{n(C_\ell)}{n(E)} \quad (9)$$

と定義し, 個体 I_i の状態数 k および次数 m は $P(k, m|\mathcal{E})$ に従う. ここで, $n(\cdot)$ は集合の個数を表す.

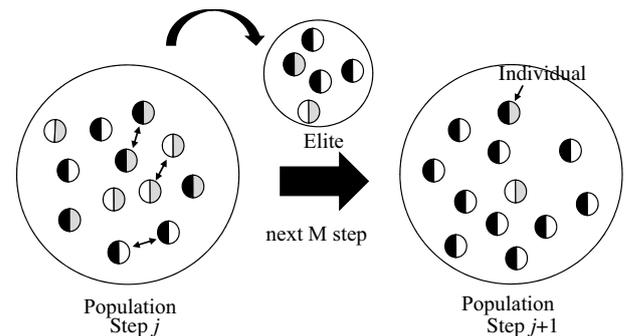


図 2 個体生成のイメージ

図 2 に個体生成のイメージを示す. 個体は 2 種類の特徴量を持つため, 個体の左領域に状態数を右領域に次数を色分けして示す. 探索の序盤では次元が様々な個体が集団内に一様に生成されるため, 大域的な解探索が行われる. 提案法では, AR-HMM の M ステップのパラメータの更新に DE を用いる. そのため, j 番目の M ステップ開始時に, 前のステップ $j - 1$ で更新したエリート個体に基づく提案個体生成分布 $P(k, m|\mathcal{E})$ によって更新されていくため, 探索が進むにつれて, 母集団内の個体の次元はひとつに収束していく. これにより, 適切な次数の候補となる特徴量を持った個体についてより局所的な探索ができる^{*2}. 以下に自己適応型 DE のアルゴリズムを示す.

^{*1} a_1 は状態 k の AR-HMM の 1 期前の自己回帰係数を表す.

^{*2} $j = 0$ の探索開始時初期個体は式 (9) の代わりに一様分布を用いて個体を生成する.

Algorithm 1 SaDE

```

1: Initialize population using  $P_{j-1}(k, m|\mathcal{E}_{j-1})$ 
2: while  $g < \text{Max Generation}$  do
3:   Mutate individuals  $\mathcal{I}$  using jDE
4:   Evaluate  $\mathcal{I}$ 
5:    $g++$ 
6: end while
7: Select Elite individuals  $\mathcal{E}$ 
8: Update  $P_j(k, m|\mathcal{E}_j)$ 
9: Update  $\Phi_j \leftarrow$  best individual
10: Return  $\Phi_j, P_{j-1}(k, m|\mathcal{E}_{j-1})$ 
    
```

次に、提案法のフレームワークについて述べる．アルゴリズムを以下に示す．

Algorithm 2 SaDE-AR-HMM

```

1: Initialize  $\mathbf{A}_0, \pi_0, \lambda_0$ 
2:    $\Phi_0 = \{\phi_1, \phi_2, \dots, \phi_K\}$ 
3:    $\Sigma_0 = \{\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2\}$ 
4:  $j = 1$ 
5: while  $Q_{j-1} < Q_i$  do
6:   Update  $\gamma_j, \xi_j \leftarrow \text{BW}(\mathcal{X}, \mathbf{A}_{j-1}, \pi_{j-1}, \Phi_{j-1})$ 
7:   Update  $\Phi_j, P_j \leftarrow \text{Adopt saDE}(\mathcal{X}, \gamma_j, \xi_j, \lambda_{j-1})$ 
8:   Update  $\mathbf{A}_j, \pi_j, \Sigma_j$ 
9:   Update  $Q_j \leftarrow \text{Evaluate}(\mathcal{X}, \mathbf{A}_j, \pi_j, \Phi_j)$ 
10:   $j++$ 
11: end while
    
```

ここで、 $\gamma(\cdot), \xi(\cdot)$ の計算には Baum-Welch アルゴリズム (BW) を用いる． \mathbf{A}, π, Σ については $\gamma(\cdot), \xi(\cdot), \Phi$ を用いて解析的に求める．また AR モデルのパラメータには依存関係があるため、パラメータを同時に更新することが求められる．そこで我々は DE の jDE とアルゴリズム SaDE を用いてパラメータの同時更新を行う．また DE の評価関数は (8) 式を使用する．

4. 実験

4.1 実験データ

提案法の有効性を検証するために 3 つのデータを用い

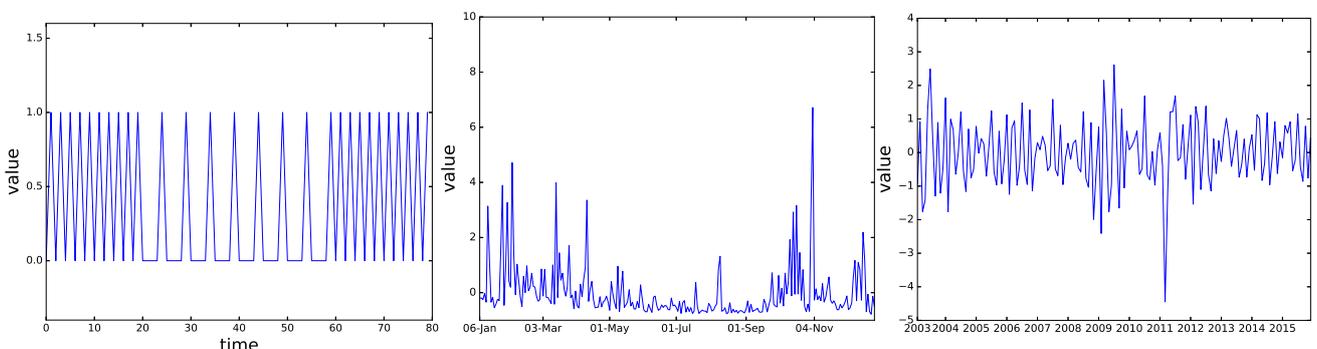
て性能を評価する．まず、モデルで推定した状態の妥当性について評価するために、予め意図して 2 つの状態を持つようなパルス信号を作成した．状態 1 では時点が $2n$ のときにのみ 1 となり、状態 2 では時点が $5n$ のときにのみ 1 となるような人工データである．次に変動が複雑な実データに対して提案モデルの有効性を確かめるため、2014/1/4~12/30 における日経平均の Realized Volatility (RV) の推定を行った．RV は株価の変動率を表しており、投資のリスク判断や、市場の状態を判断するために用いられる指標である．ここで、RV の計算には日中の 5 分毎の終値の対数収益率を用いた．最後に、周期性のあるデータから有益な状態が推定出来るのかを検証するため、日本政府観光局 (JNTO) が公開している 2003/1/1~2015/12/31 における訪日外客数データの 1 ヶ月毎の対数差分をとったデータを使用した．RV 及び訪日外客数においては、観測データを平均と分散を用いて正規化 (z-normalization) して実験に使用した．実験データを図 3 に示す．

4.2 比較法

まず、提案モデルと既存のモデルの性能を比較するため、AR モデルのパラメータ推定にユールウォーカー法 [6] を用いた AR-HMM と比較する．各パラメータの初期化は提案モデルと同様に行った．また各状態における AR モデルの次数は全て 8 とした．本論文では、このモデルを比較法 1 と呼ぶ．

次に、我々が提案した個体生成分布が AR モデルの適切な次数の推定と効率の良いパラメータの推定ができているのかを確認するために、提案個体生成分布を用いず、常に個体の生成確率が一樣な場合の DE-AR-HMM と比較する．またそのときの各パラメータは、提案モデルと同じように設定した．本論文では、このモデルを比較法 2 と呼ぶ．

最後に DE を同じ進化計算手法で高い性能と汎用性を有する手法に遺伝的アルゴリズム (GA) がある．状態数、次



(a) 人工データ (パルス)

(b) Realized Volatility

(c) 訪日外客数データ

図 3 実験データ

数および自己回帰係数に依存関係があるため、設計変数間に依存関係がある問題に強い UNDX [8] を用いた手法と性能を比較する。本論文では、この手法を GA 法と呼ぶ。

4.3 パラメータ設定

$\pi = [0.5 \ 0.5]$, $A = [0.9, 0.1], [0.1, 0.9]$ を初期値とし状態数 $K=2$, 最大次数 $M = 8$, エリート個体数 $E = 20$ とした。次に、DE および GA のパラメータ構成を示す。世代数 50, 母集団数 250 とし、DE の交叉率 0.5, スケーリングパラメータ 0.5 とした。また、GA の UNDX で用いるパラメータは、分布を生成するための個体数=10, $\alpha = 0.5$, $\beta = 0.35$, 交叉率 1.0, 突然変異率 0.0 とした。

4.4 性能評価

各実験データにおけるそれぞれのモデルの結果を図 4 から 6 に示す。上段は各モデルにおける観測データの予測値であり、下段はそのときの各状態の負担率を表している。また、各モデルの状態変数の推定には Viterbi アルゴリズムを用いた。Table1 に各モデルにおけるそれぞれの実験データから得られた AIC を示す。図 4 から 6 はある試行の結果を、Table1 は 20 回試行の平均値を示す。

図 4 はパルス信号の実験結果を示す。このテストデータは、 $n = 21 \sim 60$ を状態 1 とし、 $n = 1 \sim 20$ 及び $n = 61 \sim 80$ を状態 2 とした 2 つの状態を持っている。比較法 1 では、モデルの推定結果に多少の誤差があるが、テストデータが意図している状態を捉えることが出来た。比

較法では、モデルの推定性能は高かったが、我々が意図した状態を捉えることが出来なかった。対して提案法は、モデルの推定と状態推定を両立するできたことが分かる。

表 1 各モデルで得られた AIC

	Palse	RV	tourists
比較法 1	-28.4	602.2	693.3
比較法 2	-37.6	465.7	515.4
GA 法	-189.4	402.3	485.3
提案法	-234.9	408.6	483.5

次に、各モデルにおける日経平均の RV の推定結果を図 5 に示す。比較モデル、及び提案モデルとも RV の変動が激しい期間を状態 2 として捉えることができています。提案モデルは他の比較も出ると比べて AIC が最も低く、少ない次数で状態を捉えることが出来ているため、データの変動の本質を捉えることができています。

最後に、訪日外客数に対する各モデルの推定結果を図 6 に示す。訪日外客数のデータは状態 1 と 2 の変遷が他のデータに比べて多いが、150 日、200 日の訪問数が多い時期はどの手法も状態 1 が選択されている。また、GA 法と提案法の性能差は図 6(c) および (d) からは確認できないが、比較法 1 に比べると両手法ともデータの推定性能が高いことが分かった。

次に、AIC を比較する。表 1 に結果を示す。訪日外客数のデータに関しては、提案法は、GA 法と同程度の AIC で

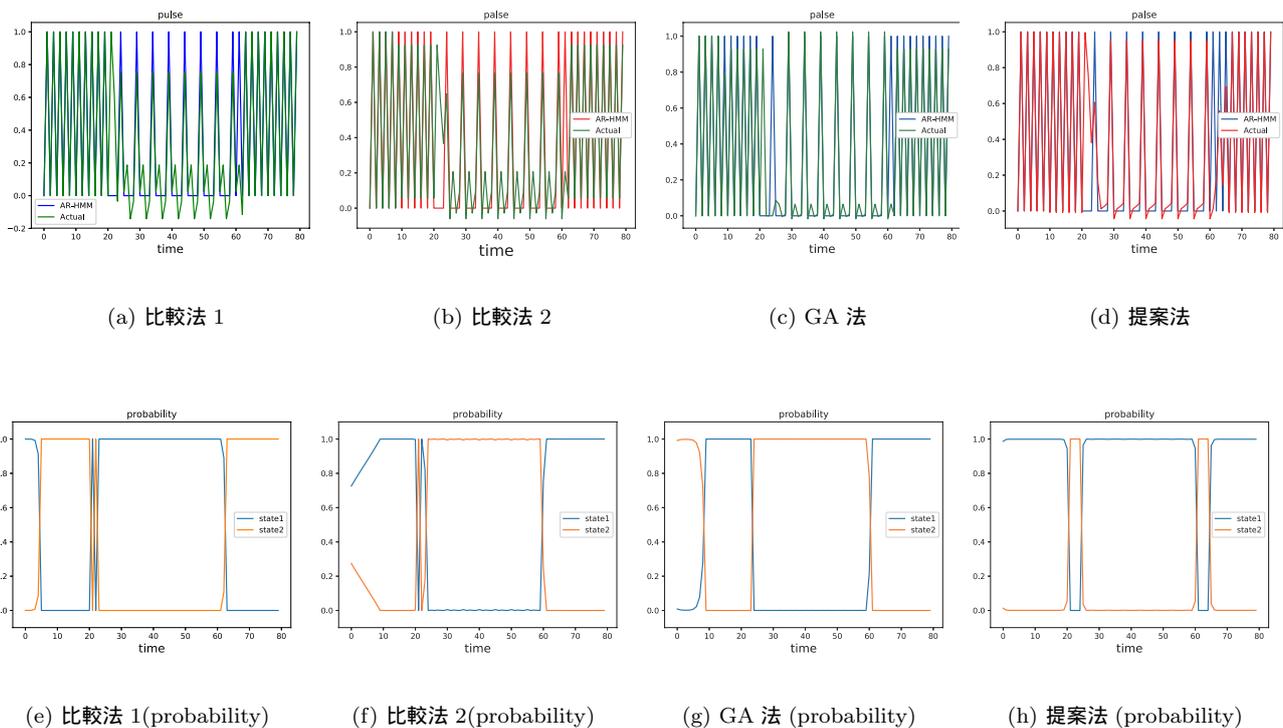
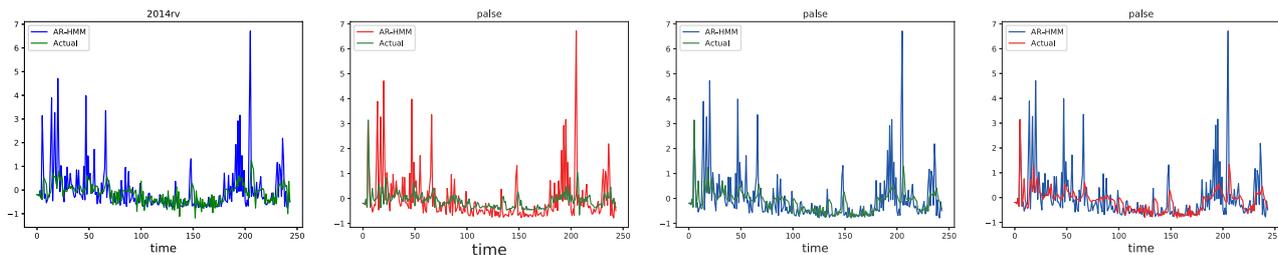
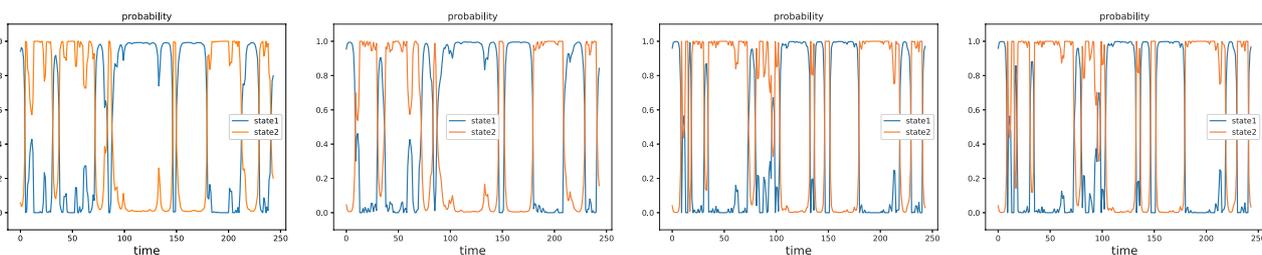


図 4 パルス信号に対してのそれぞれのモデルの予測値と状態確率

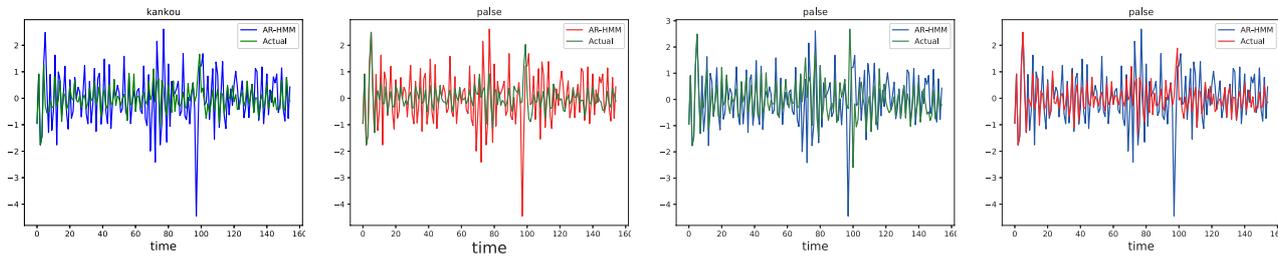


(a) 比較法 1 (b) 比較法 2 (c) GA 法 (d) 提案法

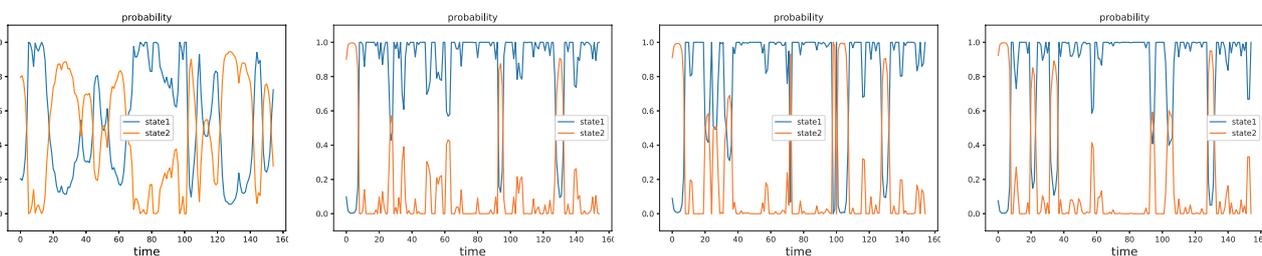


(e) 比較法 1(probability) (f) 比較法 2(probability) (g) GA 法 (probability) (h) 提案法 (probability)

図 5 RV に対してのそれぞれのモデルの予測値と状態確率



(a) 比較法 1 (b) 比較法 2 (c) GA 法 (d) 提案法



(e) 比較法 1(probability) (f) 比較法 2(probability) (g) GA 法 (probability) (h) 提案法 (probability)

図 6 訪日外客数に対してのそれぞれのモデルの予測値と状態確率

あるが、その他のデータに関しては、提案法は他の比較モデルと比べて AIC が最も低く、精度の高い推定が行われていることが分かった。これらの結果より、提案法の個体生成分布を用いた自己適応型 DE は AR-HMM の状態推定、

時系列データの推定が可能であり、従来の手法より高い性能を示すこと、多段階最適化問題の解法に提案するフレームワークは有効に働くことが分かった。

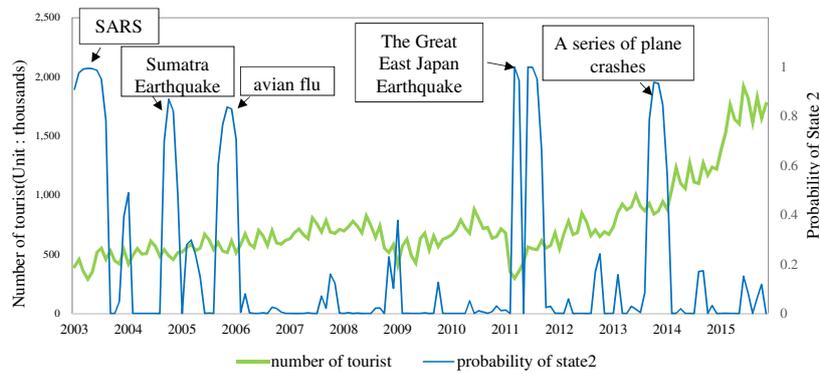


図 7 訪日外客数の状態解析

4.5 分析

GA 法と DE 法について分析を行う。実験結果より、訪日外客数データに関しては同等の性能であり、その他のデータには提案法が高い性能を示すことが分かった。Algorithm2 に示す多段階最適化問題のための提案法は、最適化の手法に様々な手法を適用可能な汎用性の高いフレームワークを有している。本論文では GA と DE を用いているが、その計算時間に関して考察を行なった*3。結果を表 2 に示す。これらは 20 回試行の平均値を示している。

表 2 計算時間の比較 (秒)

	パルス	RV	訪日外客数
GA 法	359	728	1448
提案法	254	482	951

表 2 より、データが最も少ないパルスデータが最も計算時間が短く、データがもっとも多い訪日外客数データが計算時間が長く、妥当な結果が得られていることが分かる。また、GA 法と提案法を比較すると、提案法は計算時間が短いことが分かる。GA 法では次状態の個体を生成するために確率モデルを求める必要があるが、DE 法では差分計算のみで良いために、このような差が生じたと考えられる。

次に、訪日外客数データに関して、提案法によって得られた状態の解析を行う。図 7 の緑の線は 2003/1/1~2015/12/31 における訪日外客数であり、青の線はその時の状態 2 の負担率を表している。図 7 より、状態 2 は訪日外客数が例年と比べて減少している期間を捉えていることがわかる。次に、訪問者が減っている時期に起こった出来事を調査し、提案法により抽出した状態 2 の妥当性について検証を行なった。2003 年の 1 月から 7 月にかけては、SARS と呼ばれる新型肺炎が主に中国や台湾で流行し、主となる中国人の観光客が減少したためだと考えられる。2004 年の 12 月は、スマトラ島沖地震が発生し、世界的に大きな被害を受けたためだと考えられる。2005 年の 8 月から 2006 年にかけては、中国やインドネシアをはじめとする各国で鳥インフ

*3 比較法 1 では、パラメータの組み合わせを全て比較する必要があり、提案法との計算時間の比較に意味がない。そのため、ここでは GA 法との比較を示す。

ルエンザが流行したためだと考えられる。2011 年の 3 月には東日本大震災の発生により日本が甚大な被害を受け、訪日外客数が減少したと考えられる。2013 年の 10 月から 2014 年 3 月にかけては、各国で相次いだ航空墜落事故や、政治的な理由などにより中国からの観光客が激減したためだと推測される。このように、状態 2 の負担率が高い時期に観光客数が減少する出来事が発生しており、提案法の有効性が確認できた。

5. Conclusion&Feature work

身の回りにある数多くのデータは時間による変化を伴う時系列データであり、ある期間の傾向やイベント発生による変動を分析する際など、様々な場面で時系列データ解析が必要になる。本研究では、時系列データの状態推定として用いられている AR-HMM における 2 つの問題を示し、それらの問題を解決するために SaDE-AR-HMM を提案した。提案法は 2 段階の最適化問題を同時に最適化可能な枠組みを有しており、また、適用する最適化手法は対象問題に応じて変更可能である。今回対象とした問題は複数の局所解がある問題であることが分かっているため、複数の局所解がある高次元の問題に有効な差分進化法を用いた。性能評価においては、3 種類の時系列データを用意し、データの変動が大きい、複雑なモデルが必要な問題に対しては提案モデルが従来法より有効であることを示した。また訪日外客数データにおいて、提案法から得られた時系列データの状態を解析し、有益な状態推定が行えていることを示した。今後の課題として、さら大規模な多次元時系列データに対応できるように提案モデルを改良し、データの異常区間の検知などのより現実的な実問題に取り組みたいと考えている。

参考文献

- [1] Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M.: *Time series analysis: forecasting and control*, John Wiley & Sons (2015).
- [2] Ephraim, Y., Malah, D. and Juang, B.-H.: On the application of hidden Markov models for enhancing noisy speech, *IEEE Transactions on Acoustics, Speech, and*

- Signal Processing*, Vol. 37, No. 12, pp. 1846–1856 (1989).
- [3] Ephraim, Y. and Roberts, W. J. J.: Revisiting autoregressive hidden Markov modeling of speech signals., *IEEE Signal Process. Lett.*, Vol. 12, No. 2, pp. 166–169 (online), (2005).
 - [4] Ailliot, P. and Monbet, V.: Markov-switching autoregressive models for wind time series, *Environmental Modelling & Software*, Vol. 30, pp. 92–101 (2012).
 - [5] Pandit, S. M. and Wu, S.-M.: *Time series and system analysis with applications*, Wiley (1983).
 - [6] Hamilton, J. D.: *Time series analysis*, Vol. 2, Princeton university press Princeton (1994).
 - [7] Akaike, H.: Information theory and an extension of the maximum likelihood principle, *Selected Papers of Hirotugu Akaike*, Springer, pp. 199–213 (1998).
 - [8] Ono, I. and Kobayashi, S.: A Real Coded Genetic Algorithm for Function Optimization Using Unimodal Normal Distributed Crossover., *Proceedings of the seventh international conference on genetic algorithms*, Morgan Kaufmann, pp. 246–253 (1997).