

混合ディリクレ過程モデルを利用した ARMA モデルベース時系列クラスタリング ARMA Model Based Time Series Clustering Using Dirichlet Process Mixture Models

鷲頭 祐樹†
Yuki WASHIZU

末松 伸朗†
Nobuo SUEMATSU

林 朗†
Akira HAYASHI

岩田 一貴†
Kazunori IWATA

1 はじめに

モデルベース時系列クラスタリングでは、時系列データを生み出す混合モデルを仮定し、同一要素モデルから生成されたとみなされる時系列の集合をクラスタとする。このアプローチにおいて従来は、クラスタ数を仮定した混合モデルを、EM アルゴリズムにより当てはめる手法が主であった。例えば、ARMA モデルの場合、[3] がそうである。混合ディリクレ過程 (Dirichlet Process Mixture; DPM) モデルを用いたクラスタリングは、クラスタ数も含むベイズ解析が行える手法として近年注目されている [1]。本研究では、DPM モデルを利用することで、クラスタ数についても推論が行えるような ARMA モデルベースの時系列クラスタリング法を提案する。

2 ARMA モデル

ARMA(p, q) モデルは p 次の自己回帰モデルと q 次の移動平均モデルを合成したものであり、その時系列 $\{y_t\}$ は

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t - \psi_1 u_{t-1} - \dots - \psi_q u_{t-q} \quad (1)$$

に従う [2]。ここで u_t は平均 0、分散 σ^2 の白色雑音である。このモデルのモデルパラメータは、 p 個の AR 係数 ϕ_1, \dots, ϕ_p 、 q 個の MA 係数 ψ_1, \dots, ψ_q 、そして、白色雑音の分散 σ^2 である。これらパラメータの並びを $\theta = (\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q, \sigma^2)$ と書くことにする。本研究では、定常かつ可逆な ARMA モデルに限定して考える。

3 混合ディリクレ過程モデル

本研究では、 n 本の時系列データをからなる集合 $Y = \{y_1, \dots, y_n\}$ に対する以下の様な DPM モデルを考える [1]:

$$\begin{aligned} P &\sim DP(\alpha) \\ \theta_i | P &\sim P \\ y_i | \theta_i &\sim \text{ARMA}(p, q; \theta_i) \end{aligned} \quad (2)$$

すなわち、時系列データ y_i がモデルパラメータ θ_i を持つ ARMA(p, q) から生成され、パラメータ θ_i が確率密度 P に従い、さらに、確率密度 P がディリクレ過程 $DP(\alpha)$ に従っているとい

うモデルである。ここで、 α は実数ではなく、パラメータ θ_i のとり得る空間上の測度である。

3.1 Gibbs sampler

事後分布 $p(\theta_1, \dots, \theta_n | Y)$ のような表式を得て推論を行いたいが、それは解析的には得られない。そこで MCMC (Markov chain Monte Carlo) 法の一つである Gibbs sampler によりこの事後分布に従うサンプル集合を生成し、解析を行うのが一般的なアプローチである。

Gibbs sampler は、条件付き確率分布 $p(\theta_i | \theta_{-i}, Y)$ からのサンプリングを必要とする。ここで $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$ である。モデル (2) に対するこの条件付き確率分布は

$$p(\theta_i | \theta_{-i}, Y) \propto \frac{a p(y_i)}{a + n - 1} p(\theta_i | y_i) + \sum_{j=1, j \neq i}^n \frac{p(y_i | \theta_j)}{a + n - 1} \delta(\theta_i - \theta_j) \quad (3)$$

となる。ここで、 $a = \int \alpha(\theta) d\theta$ であり、 $\delta(\cdot)$ はディラックのデルタ関数である。また、 $p(y_i | \theta_i)$ はパラメータ θ_i の ARMA(p, q) が時系列 y_i を生成する確率、 $p(y_i)$ は時系列 y_i の事前分布、そして、 $p(\theta_i | y_i)$ は時系列 y_i のみから得られたときのパラメータ θ_i の事後分布である。

式 (3) からのサンプリングが可能であれば、Gibbs sampler は以下の様な手続きとなる。

- 1: 初期値 $\theta_{1:n}^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)})$ をランダムに選ぶ。
- 2: **for** $m = 1, \dots, M$ **do**
- 3: **for** $i = 1, \dots, n$ **do**
- 4: 式 (3) に基づき $p(\theta_i^{(m)} | \theta_{-i}^{(m)}, Y)$ に従う $\theta_i^{(m)}$ を得る。
(ここで $\theta_{-i}^{(m)} = (\theta_1^{(m)}, \dots, \theta_{i-1}^{(m)}, \theta_{i+1}^{(m-1)}, \dots, \theta_n^{(m-1)})$)
- 5: **end for**
- 6: **end for**
- 7: $S = \{\theta_{1:n}^{(m)}\}_{m=1}^M$ を生成されたサンプルとして返す。

ただし、実際には burn in と呼ばれる初期のある期間のサンプルを除いたものを生成されたサンプルとして利用する。

4 提案手法

4.1 ARMA モデルの場合の問題点

対象が時系列データではなく、生成モデルに正規分布などの標準的なパラメトリックモデルを用いる場合の多くでは、式 (3) からの直接的なサンプリングが可能である。しかし、時系列モデルの場合は、一部の例外を除いて、困難に直面する。これは生成モデルに対する共役事前分布族が存在しないためである。

† 広島市立大学大学院情報科学研究科
〒731-3194 広島市安佐南区大塚東 3-4-1
Email: washizu@robotics.im.hiroshima-cu.ac.jp

より具体的には、式 (3) からのサンプリングが必要となる、 $p(y_i)$ の値の計算と事後分布 $p(\theta_i|y_i)$ からのサンプリングが問題となる。これらは、 $\bar{\alpha}(\theta) = a^{-1}\alpha(\theta)$ とすると、それぞれ、

$$p(y_i) = \int p(y_i|\theta_i)\bar{\alpha}(\theta)d\theta \quad (4)$$

$$p(\theta_i|y_i) = \frac{p(y_i|\theta_i)\bar{\alpha}(\theta_i)}{p(y_i)} \quad (5)$$

である。したがって、 $\bar{\alpha}$ を生成モデル $p(y_i|\theta_i)$ に共役なものを選ぶことができれば、 $p(y_i)$ の解析解が得られ、事後分布 $p(\theta_i|y_i)$ もその $\bar{\alpha}$ と同族の分布となり、直接的にサンプリング可能である。しかし、ARMA モデルの場合、どのような $\bar{\alpha}$ を選んでも、式 (4) の積分を解析的に解くことができず、式 (5) も直接的なサンプリングが可能な分布とはならない。

4.2 対処法

上記の二つの問題に、次のように対処する。

まず、 $p(y_i)$ は、モンテカルロ積分により求める。つまり、事前分布 $\bar{\alpha}(\theta)$ からサンプル $\theta^{(1)}, \dots, \theta^{(N)}$ を生成して、

$$p(y_i) \approx \frac{1}{N} \sum_{i=1}^N p(y_i|\theta^{(i)}) \quad (6)$$

により $p(y_i)$ の近似値を得るのである。このモンテカルロ積分は、各データ y_i に対して、最初に一度行っておけばよい。

また、 $p(\theta_i|y_i)$ からのサンプリングは、メトロポリ-ヘイスティング (MH) 法を用いて行う。MH 法は、Gibbs sampler 同様 MCMC 法の一つであるが、目的とする分布に比例する関数があれば実行でき、Gibbs sampler より適用範囲が広い。この MH 法に基づくマルコフ連鎖は各データに対して 1 つずつ用意し、サンプリングの必要が生じたときにシミュレーションを行う。

上記二つの工夫により、式 (3) に従うサンプル生成が行えるので、Gibbs sampler が実行可能となる。

5 クラスタリング

実際にデータをクラスタリングするためには、提案アルゴリズムにより得られたサンプル $S = \{\theta_{1:n}^{(m)}\}_{m=1}^M$ からクラスタを推定する手続きが必要になる。

DPM モデルの性質から、 m 番目のサンプル $\theta_1^{(m)}, \dots, \theta_n^{(m)}$ には一般に重複があり、パラメータを共有する時系列が一つのクラスタを成す。したがって、 M 個のサンプルそれぞれに対応するクラスタリングがある。

二つの時系列 y_i, y_j が似通っていれば、同一のクラスタに含まれる可能性が高いはずであるから、その事後確率 $\Pr(\theta_i = \theta_j|Y)$ をサンプルから推定し、 y_i, y_j 間の近さの尺度とする。本研究におけるクラスタリングでは、 y_i, y_j 間の距離を

$$d(y_i, y_j) = -\log \Pr(\theta_i = \theta_j|Y) \quad (7)$$

により定義し、併合型階層クラスタリングを行う。¹

6 実験

実験では人工データを用いて提案手法の有効性を検証する。人工データは ARMA(1,1) モデルに従う長さ 200 の時系列デー

タ 30 本を表 1 のように生成した。

本研究では、定常かつ可逆な ARMA モデルに限定して考える。定常かつ可逆であるためには、ARMA 係数は相互に依存するある条件を満たさなければならないので、そのままでは取り扱いが難しい。そこで、[2] に示されている ARMA 係数の定常かつ可逆である範囲から \mathbb{R}^{p+q} への変換とその逆変換を利用する。ARMA 係数 (ϕ_1, ψ_1) を変換したものを $x = (x_1, x_2)$ と表記するとき、ディリクレ過程のパラメータ α は

$$\alpha(\theta = (x, \sigma^2)) = 2N(x|0, 0.45I)\chi^{-2}(\sigma^2|3, 0.2) \quad (8)$$

として実験を行った。ここで、 0 は零ベクトル、 I は単位行列、 χ^{-2} は逆カイ二乗分布を表す。

Gibbs sampler により 2000 サンプルを生成し、最初の 1000 サンプルは burn in 期間として捨て、後の 1000 サンプルに基づいてクラスタリングを行った。得られた樹形図を図 1 に示す。ここでは、階層クラスタリングに最近傍法 (単連結法) を用いた。この樹形図を見ると、3 クラスタへ分割するとき、人工データを生成した 3 つのクラスタが正しく分割されることが分かり、提案する手法の有効性が確認できる。

表 1 人工データ

パラメータ	データ数	クラスタ番号
(0.2, 0.2, 0.2)	10	1
(0.5, 0.5, 0.2)	10	2
(0.9, 0.9, 0.2)	10	3

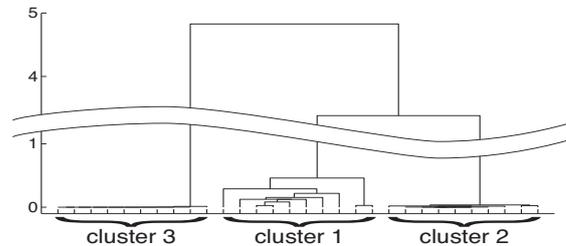


図 1 樹形図

7 まとめ

混合ディリクレモデルを利用し、時系列生成モデルに ARMA モデルを用いたクラスタ数を仮定しないモデルベース時系列クラスタリングを提案し、人工データに対する実験で有効性を確認した。ここでは、ARMA モデルを用いているので、対象として定常な時系列を考えている。しかし、4.2 で述べた対処法は他のパラメトリックな時系列モデルに対しても広く使える。

参考文献

[1] Michael D. Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588, Jun. 1995.

[2] John F. Monahan. A note on enforcing stationarity in autoregressive-moving average models. *Biometrika*, 71(2):403–404, 1984.

[3] Yimin Xiong and Dit-Yan Yeung. Mixtures of arma models for model-based time series clustering. In *Proc. of the IEEE International Conference on Data Mining*, pages 717–720, 2002.

¹ ディリクレ過程に従う確率測度は確率 1 で離散的であり、 $\Pr(\theta_i = \theta_j|Y) > 0$ となり得ることに注意されたい。