複数の時系列データに対する動的クラスタリング

Dynamic clustering for multiple time series data

数学専攻　小西 智貴
KONISHI, Tomoki

1　はじめに

IoT やデータ計測技術の発展とともに, 膨大で複雑な構造をもつ時系列データが収集されるようになった. こうした時系列データからのパターン発見を行なうために，種々の時系列クラスタリングが活用されている．Saeed et al. (2015) は, この時系列クラスタリングを時点クラスタリング, 部分列クラスタリング, 全時系列クラスタリングの 3 種類に大別している．時点クラスタリングや部分列クラスタリングは一本の長期時系列データから時点や部分列を分類する方法であり，全時系列クラスタリングは複数の時系列データを類似したグループに分類する方法である.

一方で，複数の長期時系列データに関して，時点ごとに所属クラスターや各クラスターの構造がスイッチする状況も考えられる．このような場合には，上で述べた全時系列クラスタリングなどでは表現することは不可能である．こうしたクラスター構造の動的な変化を表現できるクラスタリングを動的クラスタリングという．動的クラスタリングを行う方法としては Victhor et al. (2020) やディリクレ過程混合モデルを用いた方法がある.

本論文では，動的クラスタリングを含めて時系列データに対するクラスタリング手法についてまとめる.

2　時点クラスタリング

1 本の長期時系列データの各時点を分類する時点クラスタリングについて述べる．隠れマルコフモデルを定義し，その後隠れマルコフモデル (Bishop, 2006) による時点クラスタリングを説明する.

時点 $t = 1, \ldots, T$ に対して

$$\boldsymbol{z}_t = p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, A) = \prod_{k=1}^{K}\prod_{j=1}^{K} a_{jk}^{z_{t-1,j}z_{tk}}, \tag{2.1}$$

$$\boldsymbol{x}_t = p(\boldsymbol{x}_t|\boldsymbol{z}_t, \boldsymbol{\phi}) = \prod_{k=1}^{K} p(\boldsymbol{x}_t|\phi_k)^{z_{tk}} \tag{2.2}$$

で表されるモデルを隠れマルコフモデルという．ここで，潜在変数 $\boldsymbol{z}_t = (z_{t1}, \ldots, z_{tK})$ は K 次元の二値変数とし，$z_{tk} \in \{0, 1\}, (k = 1, \ldots, K)$ である．遷移確率 a_{jk} は $z_{t-1,j} = 1$ のとき $z_{tk} = 1$ となる確率であり，A はその遷移確率を要素にもつ行列である．また，$\boldsymbol{\phi}_t = (\phi_{t1}, \ldots, \phi_{tK})$ は観測変数 $\boldsymbol{x}_t$ の分布のパラメータである． 潜在変数 $\boldsymbol{z}_1$ は初期確率 π_k を各要素に持つ確率ベクトル $\boldsymbol{\pi}$ で表される周辺分布 $p(\boldsymbol{z}_1)$ をもち，潜在変数 $\boldsymbol{z}_1$ の条件付き分布 $p(\boldsymbol{z}_1|\boldsymbol{\pi})$ は

$$p(\boldsymbol{z}_1|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{1k}} \tag{2.3}$$

と書くことができる.

隠れマルコフモデルにおいて，潜在変数 $\boldsymbol{z}_t = (z_{t1}, \ldots, z_{tK})$ が K 次元の二値変数であることから，各時点 t の確率分布 $p(\boldsymbol{x}_t|\boldsymbol{z}_t)$ を

$$p(\boldsymbol{x}_t|\boldsymbol{z}_t, \boldsymbol{\phi}) = \sum_{k=1}^{K} z_{tk} p(\boldsymbol{x}_t|\phi_k) \tag{2.4}$$

と混合分布の形で表すことができる．式 (2.3) により，潜在変数 $\boldsymbol{z}_t$ はどの混合成分が対応する変数 $\boldsymbol{x}_t$ を生成するのかを表している．

この潜在変数をクラスターインデックスと捉えることで，各時点 $\boldsymbol{x}_t$ の所属クラスターが判明する．これが隠れマルコフモデルによる時点クラスタリングである．手順は以下の通りである．まず，クラスター数 K を定め，K を潜在変数の次元数とする．次に，パラメータ $\boldsymbol{\theta} = \{\boldsymbol{\pi}, A, \boldsymbol{\phi}\}$ を EM アルゴリズムにより推定する．このとき，最初の潜在変数 $\boldsymbol{z}_1$ は初期確率ベクトル $\boldsymbol{\pi}$ で選択され，次の潜在変数 $\boldsymbol{z}_2$ は遷移確率 $p(\boldsymbol{z}_2|\boldsymbol{z}_1)$ に従って選択される．これを繰り返すことにより，潜在変数 $Z = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_T\}$ を抽出することができ，$X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\}$ をクラスタリングすることができる．

隠れマルコフモデルでは，各時点での混合成分が独立に選択されるのではなく，過去の時点で選択された成分に依存して選択されるようになっている．つまり，隠れマルコフモデルによる時点クラスタリングは動的に各時点の所属クラスターの変化を見ることができている．

3 全時系列クラスタリング

全時系列クラスタリングは，複数の長期時系列データを類似した時系列に分類する方法であり，モデルベース，特徴ベース，形状ベースの手法，そしてその組み合わせの手法に大別される．その中でも形状ベースの手法では，時間軸を非線形に伸縮させることで，2 つの時系列の形状を可能な限り一致させ，静的データに用いられるクラスタリングアルゴリズムを時系列データそのものに対して利用する．モデルベースの手法や特徴ベースの手法では，時系列データをモデルのパラメータや低次元のベクトルに変換してからクラスタリングしている一方で，形状ベースの手法では元の時系列データに対して直接クラスタリングを行なっている．以下では，形状ベースの全時系列クラスタリングとして最も一般的に扱われる動的時間伸縮法を用いた手法 (Van and Duong, 2016) を説明する．

動的時間伸縮法（Dynamic Time Warping；DTW）とは，時系列データ同士の距離・類似度を測る手法である．DTW は時間方向の非線形な伸縮を許容し，時系列データのある 1 時点をもう片方の時系列データにおける複数時点のデータに対応づけることができる．そのため，時点数の異なる時系列データについても扱うことができる．

時点数の異なる N 個の時系列データを $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$ とする．すなわち，各時系列データを $\boldsymbol{x}_n = (x_{n1}, \cdots, x_{nT_n})$ とすると，時点数 T_n は n によって異なる．

時系列データのペア $\boldsymbol{x}_m, \boldsymbol{x}_n$ に対して，その距離を

$$\mathrm{DTW}(\boldsymbol{x}_m, \boldsymbol{x}_n) = \min\left(\sum_{k=1}^{K} |x_{mi_k} - x_{nj_k}|\right) \tag{3.1}$$

により測ることにすると，この距離を用いて $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ をクラスタリングすることができる．ここで，x_{mi_k}, x_{nj_k} はワーピングパスが通る格子点 $f_k = (i_k, j_k)$ 上の時系列データ $\boldsymbol{x}_m$，$\boldsymbol{x}_n$ の観測値である．

動的時間伸縮法によるクラスタリングでは，k-medoids 法を用いることが一般的である．クラスター数 K を定め，K 個のクラスターの中心の初期値を次の手順で決定する．まず，

$$\rho_{mn} = \frac{\mathrm{DTW}(\boldsymbol{x}_m, \boldsymbol{x}_n)}{\sum_{\ell=1}^{N} \mathrm{DTW}(\boldsymbol{x}_m, \boldsymbol{x}_\ell)}$$

とし，

$$v_n = \sum_{m=1}^{N} \rho_{mn}$$

を求める．そして，v_n が小さい K 個のオブジェクトを初期の中心とする．これは，多くのオブジェクトの近くにあるオブジェクトを中心として定めようとしていることになる．この初期値から始めて，k-medoids 法によりクラスタリングを行う．このクラスタリングによって，形状の類似した時系列データを分類することができる．

4　動的クラスタリング

全時系列クラスタリングは複数の長期時系列データを類似した時系列に分類する方法であり，各時系列が全時点を通して同じクラスターに属していると考えている．一方で，ある時点でクラスターがスイッチするような時系列データがある場合，全時系列クラスタリングでは正しくクラスタリングすることができない．そのような所属クラスターの変化に対応する時系列クラスタリングとして動的クラスタリングがある．以下では，ディリクレ過程混合モデルによる動的クラスタリングを説明する．

4.1　モデル化

ディリクレ過程混合モデルを用いて動的クラスタリングを行うために，3 つの性質を満たすモデル化が必要になる．1 つ目はクラスタリング結果が隣接時点間で高い相関をもつことである．この性質はあるデータが隣接時刻では同じクラスターに高い確率で帰属しやすいことを意味している．2 つ目はクラスタリング時間発展は絶対に時間依存し，一様でないことである．そして 3 つ目はクラスター数の時間変化を許容することである．2 つ目と 3 つ目の性質はクラスターが突発的に生成や消滅，合併や分裂を起こしながら時間発展していくことを意味している．これらの性質を満たすモデルとして

$$\boldsymbol{z}_t|\alpha \sim \mathrm{CRP}(\boldsymbol{z}_{t-1}, \alpha), \tag{4.1}$$

$$\boldsymbol{\theta}_{ti}|\beta \sim G_0(\boldsymbol{\theta}_{ti})\ (i = 1, \ldots, \infty), \tag{4.2}$$

$$\boldsymbol{x}_{tk}|\boldsymbol{z}_{tk} = w_{ti}, \boldsymbol{\theta}_{ti} \sim p(\boldsymbol{x}_t|\boldsymbol{\theta}_{ti})\ (k = 1, \ldots, N_t) \tag{4.3}$$

が考えられる．ここで，CRP は中華料理店過程である．式 (4.2) では $i = 1, \ldots, \infty$ としているが，K_t 個のクラスターを生成するには，式 (4.1) で得られたクラスター数 K_t だけ $\boldsymbol{\theta}_{ti}$ を生成すればよい．パラメータ $\boldsymbol{\theta}_{ti}$ は，各クラスターの確率分布が正規分布のときは $(\boldsymbol{\mu}_{ti}, \Sigma_{ti})$ となる．式 (4.3) の $G_0(\boldsymbol{\theta}_{ti})$ はパラメータ $\boldsymbol{\theta}_{ti}$ の事前分布を表しており，β はその事前分布のハイパーパラメータである．

4.2　潜在変数 z_t のサンプリング

潜在変数 $\boldsymbol{z}_t$ のサンプリングには中華料理店過程を用いる．潜在変数 $\boldsymbol{z}_{t-1}$ を既知とし，時点 $t \geq 2$ における z_{tk} のサンプリングを

$$\begin{aligned} p(z_{tk}|\boldsymbol{z}_{t-1,-k}, \alpha, p) &= \mathrm{CRP}(\boldsymbol{z}_{t-1,-k}, \alpha) \\ &= \begin{cases} \frac{N'_{i,t-1}}{|\boldsymbol{z}_{t-1}|-1+\alpha} & (\boldsymbol{z}_{tk} = w_{ti}), \\ \frac{\alpha}{|\boldsymbol{z}_{t-1}|-1+\alpha} & (\boldsymbol{z}_{tk} = w_{t,\mathrm{new}}) \end{cases} \end{aligned} \tag{4.4}$$

と行う．ここで，$\boldsymbol{z}_{t-1,-k}$ は時点 $t-1$ の潜在変数の中から k 番目の潜在変数を除いたものである．$w_{t,i}$ は時点 t における i 番目のクラスターであり，$w_{t,\mathrm{new}}$ は時点 t における新規クラスターである．$N'_{t-1,i}$ はクラスター $w_{t-1,i}$ に属するデータの個数であり，$|\boldsymbol{z}_{t-1}|$ は時点 $t-1$ における全てのデータの総数を表す．

式 (4.4) は α に比例した確率で新しいクラスターを生成する．すなわちパラメータ α によってクラスター数が変動する．潜在変数は一定の確率 p で変化するものとして考えるため，確率 p で $\boldsymbol{z}_{tk} \sim \mathrm{CRP}(\boldsymbol{z}_{t-1}, \alpha)$，確率 $1-p$ で $\boldsymbol{z}_{tk} = \boldsymbol{z}_{t-1,k}$ とする．

観測値が得られた際，潜在変数 $\boldsymbol{z}_t$ を推定するには，ギブスサンプリングを用いて潜在変数の事後確率を求める．便宜上，$\boldsymbol{z}_t$ から $\boldsymbol{z}_{tk}$ を除いた集合を $\boldsymbol{z}_{t,-k}$ とし，$\boldsymbol{x}_t$ から $\boldsymbol{x}_{tk}$ を除いた集合を $\boldsymbol{x}_{t,-k}$ と定義する．このとき，事後確率 $p(\boldsymbol{z}_{tk} = w_i|\boldsymbol{x}_{tk}, \boldsymbol{x}_{t-1,-k}, \boldsymbol{z}_{t-1,-k}, \boldsymbol{\theta}_t)$, $(i = 1, 2, \ldots)$ は

$$\begin{aligned} p(\boldsymbol{z}_{tk} = w_{ti}|\boldsymbol{x}_{tk}, \boldsymbol{x}_{t-1,-k}, \boldsymbol{z}_{t-1,-k}, \boldsymbol{\theta}_t) &= \frac{p(\boldsymbol{x}_{tk}, \boldsymbol{z}_{tk} = w_{1i}|\boldsymbol{x}_{t-1,-k}, \boldsymbol{z}_{t-1,-k}, \boldsymbol{\theta}_t)}{p(\boldsymbol{x}_{tk}|\boldsymbol{x}_{t-1,-k}, \boldsymbol{z}_{t-1,-k}, \boldsymbol{\theta}_t)} \\ &\propto p(\boldsymbol{x}_{tk}|\boldsymbol{z}_{tk} = w_{ti}, \boldsymbol{\theta}_t)p(\boldsymbol{z}_{tk} = w_{ti}|\boldsymbol{z}_{t-1,-k}) \end{aligned} \tag{4.5}$$

となる．式 (4.5) の $p(\boldsymbol{x}_{tk}|\boldsymbol{z}_{tk}=w_{ti},\boldsymbol{\theta}_t)$ は

$$p(\boldsymbol{x}_{tk}|\boldsymbol{z}_{tk}=w_{ti},\boldsymbol{\theta}_t)=\begin{cases}p(\boldsymbol{x}_{tk}|\boldsymbol{\theta}_{ti}) & (\boldsymbol{z}_{tk}=w_{ti}),\\ \int p(\boldsymbol{x}_{tk}|\boldsymbol{\theta}_{t,\mathrm{new}})G_0(\boldsymbol{\theta}_{t,\mathrm{new}})d\boldsymbol{\theta}_{t,\mathrm{new}} & (\boldsymbol{z}_{tk}=w_{t,\mathrm{new}})\end{cases} \tag{4.6}$$

とする．式 (4.6) の下段の積分は $G_0(\boldsymbol{\theta}_{t,\mathrm{new}})$ が共役事前分布であれば解析的に求めることができ，そうでなければメトロポリス法などによるサンプリングを行う．

式 (4.5) の $p(\boldsymbol{z}_{tk}=w_{ti}|\boldsymbol{z}_{t-1,-k})$ は式 (4.4) を用いる．データ $\boldsymbol{x}_{tk}$ の所属クラスターを更新するには，$\boldsymbol{z}_{tk}=w_{ti}\,(i=1,2,\dots)$ と $\boldsymbol{z}_{tk}=w_{t,\mathrm{new}}$ となる事後確率をそれぞれ求める．そしてその結果を用いて，$\boldsymbol{z}_{tk}$ の値を確率的に決定する．この処理を k の値を変えながら続けることで，次々に $\boldsymbol{z}_{tk}$ の値を更新していく．

4.3 パラメータ θ_t の更新

$\boldsymbol{z}_{t1},\cdots,\boldsymbol{z}_{tN_t}$ の値が決まれば，事前分布 $G_0(\boldsymbol{\theta}_{ti})$ とクラスター w_{ti} に所属する $\boldsymbol{x}_{tk}$ に対する尤度 $p\Big(\{\boldsymbol{x}_{tk}\in w_{ti}\}|\boldsymbol{\theta}_{ti}\Big)$ を用いて，パラメータ $\boldsymbol{\theta}_{ti}$ の事後分布は

$$p\Big(\boldsymbol{\theta}_{ti}|\{\boldsymbol{x}_{tk}\in w_{ti}\}\Big)=\frac{G_0(\boldsymbol{\theta}_{ti})\prod_{\boldsymbol{x}_{tk}\in w_{ti}}p(\boldsymbol{x}_{tk}|\boldsymbol{\theta}_{ti})}{\int G_0(\boldsymbol{\theta}_{ti})\prod_{\boldsymbol{x}_{tk}\in w_{ti}}p(\boldsymbol{x}_{tk}|\boldsymbol{\theta}_{ti})d\boldsymbol{\theta}_{ti}} \tag{4.7}$$

と表すことができる．ここで $\prod_{\boldsymbol{x}_{tk}\in w_{ti}}p(\boldsymbol{x}_{tk}|\boldsymbol{\theta}_{ti})$ はサンプリングにより次々に決定された $\boldsymbol{z}_{t1},\dots,\boldsymbol{z}_{tN_t}$ のうち，$\boldsymbol{z}_{tk}=w_{ti}$ となる $\boldsymbol{x}_{tk}$，すなわち $\boldsymbol{x}_{tk}\in w_{ti}$ となる $\boldsymbol{x}_{tk}$ のみを取り出して積をとることを意味している．この事後分布に基づいて K_t 個のクラスターのパラメータである $\boldsymbol{\theta}_{ti}\,(i=1,2,\dots,K_t)$ を確率的に更新する．

5 おわりに

本論文では，Saeed et al. (2015) が扱っていない動的クラスタリングを含めて時系列データに対するクラスタリングを整理した．時系列データに対するクラスタリングは確率分布によるものや符号長など情報理論を用いたものなど，複数のアプローチが見られる．それらを統一的に扱った文献は見られないようであり，本論文は一定の意義があると考えられる．

複数の時系列データに対する全時系列クラスタリングが一般的に用いられるが，時間によって所属クラスターやクラスター構造が変化する場合には動的クラスタリングが有効である．しかしながら，ディリクレ過程混合モデルを用いた場合にはクラスター数がパラメータ α や p に依存し，どの程度のクラスター数となるかがコントロールしにくく，さらには時点によらず一定である．この点を解決したモデルの開発が今後の課題である．

参考文献

[1] Bishop, C. M. (2006). Pattern recognition and machine learning, Springer.

[2] Saeed, A. *et al.* (2015). Time-series clustering - A decade review, *Information Systems*, pp.16-38.

[3] Van, T. H. and Duong, T. A. (2016). An efficient implementation of anytime K-medoids clustering for time series under dynamic time warping, *in Proc. 7th Symp. Inf. Commun. Technol.*, pp.22-29.

[4] Victhor,S.S. *et al.*(2020). Dynamic clustering of time series data, *arXiv:2002.01890.*