

電力需要予測に対する モデルベース時系列クラスタリングの応用

白瀧 豪^{1,a)} 今井 貴史¹ 河本 薫¹ 國政 秀太郎²

概要: 電力事業において、電力は貯めておくことが難しいエネルギーであるため、電力需要を予測して需要と供給のバランスを整えることが必要になる。しかし、事業開始直後などは顧客数が少ないため、全顧客の需要合計を一括して予測しても、十分な精度がえられなかった。そこで、本研究では予測精度を上げることを目的とし、その達成のためにクラスタリングとクラスターごとのモデル推定を同時に行うことができる混合 ARMA モデルを用いた時系列クラスタリングを適用した。その結果、全顧客を一つの ARMA モデルで説明する場合や顧客ごとに ARMA モデルで説明する場合のどちらよりも予測精度を向上させることが確認できた。

キーワード: 電力需要予測, 時系列解析, 時系列クラスタリング

Application of Model-Based Time Series Clustering to Electricity Demand Forecasting

GO SHIRATAKI^{1,a)} TAKASHI IMAI¹ KAORU KAWAMOTO¹ SHUTARO KUNIMASA²

Abstract: In the electric power business, since electricity is an energy source that is difficult to store, it is necessary to forecast electricity demand to maintain a balance between supply and demand. However, due to the small number of customers immediately after the start of the business, it is often not possible to obtain sufficient accuracy by directly forecasting the total demand of all customers. In this study, in order to improve the forecasting accuracy, we applied a model-based method for time series clustering, which can simultaneously perform clustering and model estimation for each cluster. The results confirm that this method provides better prediction accuracy than both the case where one model is defined for all customers and the case where a model is defined for each customer.

Keywords: Electricity Demand Forecasting, Time Series Analysis, Time Series Clustering

1. 背景

電力事業において、電力は貯めておくことが難しいエネルギーであるため、時刻別に電力需要を予測して需要と供給のバランスを整えることが必要になる。顧客数が充分に多い場合は、顧客ごとの需要パターンに違いはあっても、

全顧客で合計すればパターンは平準化され、全体需要は十分な精度で予測できる。しかしながら、事業開始時など顧客数が少ない場合は、全顧客で合計しても顧客ごとの需要パターンの影響は無視できないため、全体需要は十分な精度で予測できない。そこで、需要パターンが類似している顧客群をクラスター化し、クラスターごとにそれに属する顧客単位の需要を予測するモデルを作り、それを合計して全体需要を予測することで予測精度を向上させることを考えた。具体的には、クラスタリングとクラスターごとのモデル推定を同時に行うことができる混合 ARMA モデルを用いた時系列クラスタリング [3] を適用する。

¹ 滋賀大学
Shiga University

² 大阪ガス(株) 情報通信部
Information/Communication Systems Dept., Osaka Gas Co., Ltd.

a) g.shirataki4@gmail.com

2. 適用手法

2.1 時系列クラスタリングの概要

近年、ストレージの増加や計算性能の向上により、生物学、ファイナンス、気象など様々な分野でデータが大量に蓄積されてきている。そして、それらのデータの多くは時系列データとして扱うことができ、時系列クラスタリングはデータマイニング研究の対象として重要性を増してきている [1][2]。

時系列クラスタリングは大きく3つに分類することができる。

全体時系列クラスタリング

個別の時系列データに対してクラスタリングを行う。

部分時系列クラスタリング

単一の時系列データから部分系列を複数取り出してクラスタリングを行う。

時間点クラスタリング

時系列データの時間点ごとにクラスタリングを行う。

また、全体時系列クラスタリングはさらに3つのアプローチに分けられる。

形状ベースアプローチ

生の時系列データをそのまま従来のクラスタリング手法に適用する。

特徴ベースアプローチ

生の時系列を低次元の特徴ベクトルに変換し、抽出した特徴ベクトルに対して従来のクラスタリング手法を適用する。

モデルベースアプローチ

クラスターごとにモデルを仮定し、それらのモデルをデータに適合させることでクラスタリングを行う。

2.2 混合 ARMA モデルを用いた時系列クラスタリング手法

本研究の対象としては、電力使用の傾向により顧客をクラスタリングするという考えから全体時系列クラスタリングが最適であると考えた。またクラスタリングとモデルの予測精度が向上することを期待してモデルベースの時系列クラスタリングを採用する。特に今回は、文献 [3] の手法を適用する。

文献 [3] の手法ではシステムのモデルとして混合 ARMA モデルを仮定する。混合 ARMA モデルとは、複数の異なる自己回帰移動平均 (Autoregressive-Moving-Average: ARMA) モデルを考え、それらの一つ一つに時系列が従う確率を混合要素とした混合モデルである。混合 ARMA モデルを構成する個々の ARMA モデルがクラスターに対応すると考えられる。この混合 ARMA モデルに対して、各 ARMA モデルのパラメータと混合係数の学習を期待値最大化 (Expectation-Maximization: EM) アルゴリズム

で推定を行い、ベイズ情報量規準 (Bayesian Information Criterion: BIC) を用いてクラスター数の選択と同時にモデル選択を行う。

2.2.1 一般的な ARMA モデル

自己回帰次数 p 、移動平均次数 q をもつ定常 ARMA モデルを $ARMA(p, q)$ と表す。時系列データ $\mathbf{x} = \{x_t\}_{t=1}^n$ に対して $ARMA(p, q)$ を適用すると以下のように表せる。

$$x_t = \phi_0 + \sum_{j=1}^p \phi_j x_{t-j} + \sum_{j=1}^q \theta_j e_{t-j} + e_t \quad (1)$$

ただし、 $\{e_t\}_{t=1}^n$ は $e_t \sim N(0, \sigma^2)$ かつホワイトノイズであり、 $\Phi = \{\phi_0, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2\}$ は AR 成分と MA 成分およびホワイトガウスノイズの分散の集合である。

時系列 $\mathbf{x} = \{x_t\}_{t=1}^n$ の条件付き対数尤度関数は以下のように表せる [4]。

$$\log P(\mathbf{x} | \Phi) = -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^n e_t^2 \quad (2)$$

ここで、 e_t は式 (1) から以下のように計算することができる。

$$e_t = x_t - \phi_0 - \sum_{j=1}^p \phi_j x_{t-j} - \sum_{j=1}^q \theta_j e_{t-j} \quad (3)$$

2.2.2 混合 ARMA モデル

一般的な ARMA モデルを混合モデルに拡張したモデルを混合 ARMA モデルと呼ぶ。

M 個の ARMA モデルを仮定する。このときクラスターごとにモデルを定義するため、クラスター数が M 個であることも仮定している。また、各クラスターを $\omega_1, \dots, \omega_M$ とする。

クラスター k のモデルパラメータを Φ_k 、尤度を $P(\mathbf{x} | \omega_k, \Phi_k)$ 、クラスター k の事前分布を $P(\omega_k)$ とすると、混合 ARMA モデルの尤度関数は以下のように表せる。

$$P(\mathbf{x} | \Theta) = \sum_{k=1}^M P(\mathbf{x} | \omega_k, \Phi_k) P(\omega_k)$$

ただし、 $\Theta = \{\Phi_1, \dots, \Phi_M, P(\omega_1), \dots, P(\omega_M)\}$ であり、これが混合 ARMA モデルのモデルパラメータとなる。

次に、 N 個の時系列データの集合を $D = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ とする。各時系列データの独立性を仮定すると、集合 D の尤度は以下ようになる。

$$P(D | \Theta) = \prod_{i=1}^N P(\mathbf{x}_i | \Theta) \quad (4)$$

この尤度の下で、パラメータ Θ の最尤推定量 $\hat{\Theta}$ は以下のようになる。

$$\hat{\Theta} = \arg \max_{\Theta} (P(D | \Theta))$$

この最尤推定量 $\hat{\Theta}$ を EM アルゴリズムを用いて推定する。

2.2.3 EM アルゴリズム

混合 ARMA モデルのパラメータ推定をするために各時系列データがどのクラスターに属するかの情報が必要になる。そこで時系列データ i が属するクラスターの番号を z_i として潜在変数 $Z = \{z_1, \dots, z_N\}$ を導入し、完全データ (D, Z) とする。 z_i は時系列データ i が属するクラスターの番号を表しているため、以下を満たす。

$$P(\mathbf{x}_i, z_i | \omega_k, \phi_k) = \begin{cases} P(\mathbf{x}_i | \omega_k, \phi_k) & \text{if } k = z_i \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

集合 D の尤度は式 (4) であった。式 (5) を用いて、対数尤度 l は以下のように計算できる。

$$\begin{aligned} l(\Theta; D, Z) &= \sum_{i=1}^N \log P(\mathbf{x}_i, z_i | \Theta) \\ &= \sum_{i=1}^N \log \left(\sum_{k=1}^M P(\mathbf{x}_i, z_i | \omega_k, \phi_k P(\omega_k)) \right) \\ &= \sum_{i=1}^N \log (P(\mathbf{x}_i | \omega_{z_i}, \phi_{z_i}) P(\omega_{z_i})) \\ &= \sum_{i=1}^N \log P(\mathbf{x}_i | \omega_{z_i}, \phi_{z_i}) + \sum_{i=1}^N \log P(\omega_{z_i}) \end{aligned} \quad (6)$$

次に、 Z に関する対数尤度の期待値 Q が最大となるようにパラメータの更新を行う。そのため t 回目時点のパラメータを $\Theta(t)$ と表すと、 Z に関する対数尤度の期待値 Q は以下のように計算できる。

$$\begin{aligned} Q(\Theta | \Theta(t)) &= E \left[\sum_{i=1}^N \log P(\mathbf{x}_i | \omega_{z_i}, \phi_{z_i}), \Theta(t) \right] \\ &\quad + E \left[\sum_{i=1}^N P(\omega_{z_i}), \Theta(t) \right] \\ &= \sum_{i=1}^N \sum_{z_i=1}^M P(\omega_{z_i} | \mathbf{x}_i, \Theta(t)) \log P(\mathbf{x}_i | \omega_{z_i}, \phi_{z_i}) \\ &\quad + \sum_{i=1}^N \sum_{z_i=1}^M P(\omega_{z_i} | \mathbf{x}_i, \Theta(t)) \log P(\omega_{z_i}) \\ &= \sum_{i=1}^N \sum_{k=1}^M P(\omega_k | \mathbf{x}_i, \Theta(t)) \log P(\mathbf{x}_i | \omega_k, \phi_k) \\ &\quad + \sum_{i=1}^N \sum_{k=1}^M P(\omega_k | \mathbf{x}_i, \Theta(t)) \log P(\omega_k) \end{aligned} \quad (7)$$

また、

$$P(\omega_k | \mathbf{x}_i, \Theta(t)) = \frac{P(\mathbf{x}_i | \omega_k, \phi_k) P(\omega_k)}{\sum_{u=1}^M P(\mathbf{x}_i | \omega_u, \phi_u) P(\omega_u)} \quad (8)$$

したがって EM アルゴリズムでは、まず式 (7) から関数 Q

Algorithm 1 EM アルゴリズム

Input: $\Theta(1), \epsilon$

- 1: $t \leftarrow 0$
- 2: **while** $l(\Theta(t+1); D) - l(\Theta; D) > \epsilon$ **do**
- 3: $t \leftarrow t + 1$
- 4: E ステップ: $P(\omega_k | \mathbf{x}_i, \Theta(t))$ と $Q(\Theta | \Theta(t))$ を算出
- 5: M ステップ: $\Theta(t+1) \leftarrow \arg \max_{\Theta} Q(\Theta | \Theta(t))$
- 6: **end while**
- 7: **return** $\Theta(t+1)$

図 1 EM アルゴリズム

を計算し、式 (8) から $P(\omega_k | \mathbf{x}_i, \Theta(t))$ を計算する (E ステップ)。そして次に、計算した関数 Q が最大となるようなパラメータ $\hat{\Theta}$ を計算し、 $\Theta(t+1) = \hat{\Theta}$ とする (M ステップ)。このときもし尤度が収束条件を満たした場合、その時点のパラメータ $\Theta(t+1)$ を推定されたパラメータとする。すなわち $l(\Theta(t+1); D) - l(\Theta; D) < \epsilon$ を満たした時点のパラメータ $\Theta(t+1)$ を推定されたパラメータとする。収束条件が満たされなかった場合は E ステップに戻る。アルゴリズムをまとめたものが図 1 である。

2.2.4 初期値の設定

EM アルゴリズムは局所最適解に陥りやすいアルゴリズムと知られている。よって、初期パラメータに大きく依存することが考えられる。そのような中で比較的妥当な解を得るため、個々の時系列から見積もられたパラメータをヒントに初期パラメータを決定する。すなわち、各時系列データに対して ARMA モデルの推定し、係数ベクトルを特徴量ベクトルとして k-means++法を適用する。適用後の各クラスターの重心点を EM アルゴリズムの初期パラメータ $\Theta(1)$ とする。

2.2.5 モデル選択

モデル選択には BIC を用いる。BIC は一般的に以下の式で表される。

$$\text{BIC} = \log P(D, \hat{\Theta}) - \frac{1}{2} V \log N$$

ただし、 $\hat{\Theta}$ は推定されたモデルのパラメータであり、 V はモデルのパラメータ数である。

BIC では第一項が尤度になっておりモデルの当てはまりの良さを評価している。また、第二項で罰則項を設定している。したがって、BIC が高いほどモデルの評価が高いことを示す。

混合 ARMA モデルの場合の BIC は以下のように書き表すことができる。

$$\begin{aligned} \text{BIC} &= \sum_{i=1}^N \log \left(\sum_{k=1}^M P(\mathbf{x}_i | \omega_k, \phi_k P(\omega_k)) \right) \\ &\quad - \frac{1}{2} (Mv + M - 1) \log N \end{aligned} \quad (9)$$

ただし、 M はクラスター数であり、 v はクラスターごとの ARMA モデルのパラメータ数である。

モデル選択までの流れをまとめると以下ようになる。

- (1) 適用するモデル (ARMA(p, q) の p と q) とクラスターの最大数を決める。
- (2) クラスター数を 2 から (1) で決めたクラスターの最大数までのアイドルで設定し、それぞれに対して EM アルゴリズムを実行する。
- (3) 各クラスター数に対して BIC を保存しておく。
- (4) 最大の BIC を持つクラスター数とモデルを選択する。

3. 使用データの詳細

3.1 データ内容

本研究では 100 ユーザーを対象に 1 時間ごとに計測した電力消費量データを 1 年間分使用した。つまり時系列データ \mathbf{X} は、ユーザー i の電力消費時系列を \mathbf{x}_i とすると $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_{100}\}$ と表すことができる。また、任意の時系列データ $\mathbf{x}_i \in \mathbf{X}$ は、時刻 t におけるユーザー i の電力消費量を $x_{i,t}$ とすると $\mathbf{x}_i = \{x_{i,1}, \dots, x_{i,8760}\}$ と表すことができる。

3.2 前処理

電力需要のデータは週による周期があると考えられるが ARMA モデルではそれを考慮することができないため、前処理として 1 週間 (7 日間) の差分をとる。差分をとった時系列データを $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_{100}\}$ とすると任意の $\mathbf{y}_i = \{y_{i,1}, \dots, y_{i,8592}\} \in \mathbf{Y}$ は以下のように表せる。

$$y_{i,t-24*7} = x_{i,t} - x_{i,t-24*7}$$

ただし、 $t \in \{169, \dots, 8760\}$ とし、初週のデータは差分をとることができないため使用しない。前処理をしたデータの推移を可視化すると図 2 のようになる。全時系列データに対して有意水準 1% で単位根検定を行い、定常性を確認ができています。

また任意の $j \in \{1, \dots, 8592\}$ に対して、 $y_{i,j} = 0$ となるような時系列データ \mathbf{y}_i は本研究では対象外とする。

3.3 予測方法と評価指標

予測性能を検証するためのデータは最後の 1 ヶ月分 (31 日分) とし、混合 ARMA モデルの混合要素として AR(2) モデルを使用する。AR(2) モデルの式は以下のように表せる。

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \quad (10)$$

ただし、 $\{\epsilon_t\}_{t=1}^{8592}$ は $\epsilon_t \sim N(0, \sigma^2)$ かつホワイトノイズである。

また最大クラスター数 M は 10 とする。

予測は常に 4 時間先を行うこととする。本研究でのデー

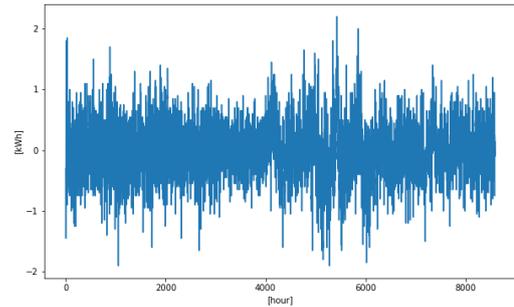


図 2 典型的な前処理後データの推移

Fig. 2 Time course of y in typical preprocessed data.

表 1 クラスターごとのユーザー数

Table 1 Number of users for each cluster.			
クラスター番号	ユーザー数	MAE[kWh]	MAPE
0	14	1.764	0.175
1	1	2.920	1.077
2	3	1.141	0.340
3	80	4.422	0.163

タは 1 時間おきにデータが取れているため 4 時点先の予測と同義になる。

本研究での予測の評価指標は平均絶対誤差 (Mean-Absolute-Error: MAE) を用いる。任意のユーザー i の予測値を $\{\hat{y}_{i,t}\}_{t=1}^{720}$ とし、MAE を算出する対象となるユーザー数を n としたときの MAE を以下のように定義する。

$$\text{MAE} = \frac{1}{720} \sum_{t=1}^{720} \left| \sum_{i=1}^n (y_{i,t} - \hat{y}_{i,t}) \right|$$

本研究での関心はユーザー全体での誤差にあるため、ユーザー全体での総和をとった後に絶対値をとっている。

またクラスター間での誤差を比較する場合にはクラスター間でのデータ量に差があるため平均絶対誤差率 (Mean-Absolute-Percentage-Error: MAPE) を使用する。MAPE は差分をとる前のスケールに戻してから算出する。すなわち時刻ごとに 1 週間前の実測値との差分の予測値を計算した後、元のスケールに戻した予測時系列 $\{\hat{x}_{i,t}\}$ を $\hat{x}_{i,t} = x_{i,t-24*7} + \hat{y}_{i,t}$ と算出する。このとき MAPE を以下のように定義する。

$$\text{MAPE} = \frac{\sum_{t=1}^{720} \left| \sum_{i=1}^n (x_{i,t} - \hat{x}_{i,t}) \right|}{\sum_{t=1}^{720} \sum_{i=1}^n x_{i,t}}$$

4. 結果と考察

本研究の使用データを混合 ARMA モデルに適用した結果、クラスター数 4 が選択された。各クラスターに属するユーザー数、MAE、MAPE の結果を表 1 にまとめた。

この結果を以下の二つの手法と比較する。

共通 AR モデル

全時系列データの各時点における平均をとったデータ

表 2 各手法ごとの結果

手法名	MAE[kWh]
混合 ARMA モデル	4.047
共通 AR モデル	5.535
個別 AR モデル	5.238

で AR(2) モデルを推定し予測を行う。

個別 AR モデル

それぞれの時系列データに対して、AR(2) モデルを推定し予測を行う。

混合 ARMA モデルを適用した結果とこれらのモデルを適用した結果は表 2 のようになる。表 2 の結果から混合 ARMA モデルが最も良い予測精度を示したことがわかる。

表 1 の結果より、クラスターごとのユーザー数に偏りがあることがわかる。特にクラスター 1 は 1 ユーザーのみのクラスターを形成している。また、各クラスターの MAPE に注目すると、クラスター 1 が他のクラスターと比較して高い数値を出している。つまり、予測精度が悪いクラスターになっていることがわかる。クラスター 1 に属するユーザーの推移を可視化したものが図 3 である。図 2 と比較すると縦軸のブレ幅が大きい、つまり電力の使用量が不規則かつ大幅に変わっていることがわかる。

次に、各クラスターのモデルパラメータを表 3 にまとめた。モデルパラメータの記号については式 (10) を参照する。クラスター 2 について、表 3 のホワイトガウスノイズ σ^2 や表 2 の MAPE に注目するとクラスター 1 に続いて高い数値となった。またこのクラスターに属するユーザー数は 3 と少ないことから挙動が乱雑なユーザーのクラスターではないかと考える。クラスター 1 やクラスター 2 と比較してクラスター 3 は、ホワイトガウスノイズ σ^2 と MAPE のどちらも低い数値であった、つまり予測精度が良いことがわかる。またこのクラスターに属するユーザー数は 80 であり、大多数のユーザーに対しては 1 つのモデルで十分な予測ができています。

これらの考察から [3] の時系列クラスタリング手法には異常データを独立したクラスターとして除外する働きがあると考えられる。つまり異常データのクラスターに対して個別のモデルが定義されたことによって、全体として予測精度の向上に繋がったと考えられる。

5. 結論

本研究では、電力需要予測に対して混合 ARMA モデルを用いた時系列クラスタリングの適用を試みた。全顧客に 1 つの ARMA モデルを適用した場合や各顧客それぞれに対して ARMA モデルを適用した場合と比較して、クラスターごとに ARMA モデルを適用する混合 ARMA モデルの方が予測精度が向上する結果が得られた。本研究では

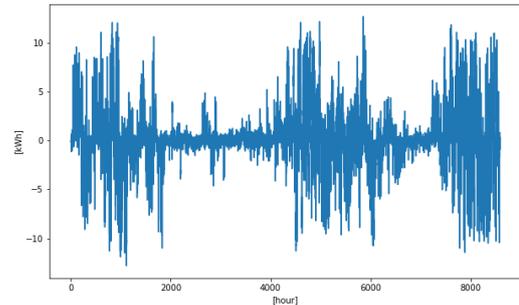


図 3 異常データの推移

Fig. 3 Time course of y in abnormal data

表 3 クラスターごとのモデルパラメータ

Table 3 Model parameters for each cluster.

クラスター番号	ϕ_0	ϕ_1	ϕ_2	σ^2
0	-0.0028	0.6544	-0.0049	0.1733
1	0.0223	0.7181	0.0073	0.3948
2	-0.0016	0.6787	-0.0683	0.3778
3	-0.0012	0.5217	0.0301	0.1308

ARMA モデルという特定の手法だけを用いたが、このことを拡大解釈すると、モデルベースの時系列クラスタリングを使用することは精度の高い予測モデルを構築する上でも有用であるということを示している。

最後に本研究の今後の展望について記す。今回適用した手法では、EM アルゴリズムの局所最適解に陥りやすい性質は初期値設定のみでは解決できず、モデルの安定性がない。また電力需要という人の予測する上で ARMA モデルではモデルの柔軟性にかけるという問題点があると考えられる。この問題に関連して、文献 [5] ではパラメータの推定に変分ベイズを用いる手法が提案されている。変分ベイズを用いることで EM アルゴリズムの局所最適解に陥りやすいという問題の軽減とモデルの安定性を期待することができる。さらには、ARMA モデルよりもより柔軟なモデリングが可能である状態空間モデルの時系列クラスタリングの手法が文献 [6] と [7] で提案されている。これらの手法を適用することによってさらなる予測精度の向上が見込まれる。

参考文献

- [1] Liao, W.T.: Clustering of time series data – a survey, *Pattern Recognition*, Vol.38, No.11, pp.1857–1874 (2005).
- [2] Aghabozorgi, S., Shirkhorshidi, S.A. and Wah, Y.T.: Time-series clustering – A decade review, *Information Systems*, Vol.53, pp.16–38 (2015).
- [3] Xiong, Y. and Yeung, D.-Y.: Time series clustering with ARMA mixtures, *Pattern Recognition*, Vol.37, No.8, pp.1675–1689 (2004).
- [4] Box, E.P.G., Jenkins, M.G. and Reinsel, C.G.: *Time Series Analysis: Forecasting and Control*, 3rd edition, (1994).

- [5] Kini, V. and Sekhar, C.: Bayesian mixture of AR models for time series clustering, *Pattern Analysis & Applications*, Vol.16, No.2, pp.179–200 (2011).
- [6] Wilcox, A.B. and Hamano, F.: Kalman’s Expanding Influence in the Econometrics Discipline, *IFAC-PapersOnLine*, Vol.50, No.1, pp.637–644 (2017).
- [7] Wilcox, A.B. and Hamano, F.: Model-Based Clustering of Time Series Based on State Space Generative Models, *Lecture Notes in Electrical Engineering*, Vol.465, pp.447–456 (2018).