LTE 環境における応答遅延特性の時系列モデリングを用いた分析

令和2年4月2日

-進捗報告-

- ・原稿執筆を行っております
- ·SINET のウェブサーバを対象とした結果と比較を行いました
- ・最適クラスタ数を求めました
- ・異なるクラスタリング手法において形成されるクラスタ間の類似度を調べています

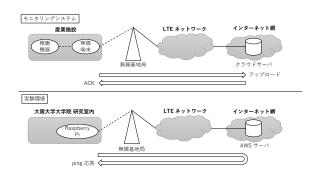


図 1: モニタリングシステムと実験環境の対応図

1 計測実験の設定

本報告では図 1 に示す構成で計測実験を行った。モニタリングシステムにおける無線端末としては LTE モジュールとして Quectel 社製 EC21-J を搭載した Raspberry Pi を用いた。また,LTE 回線としては IIJ モバイル社のサービスタイプ D 定額プランライト(いちねん プリペイド)を用いた。IIJ モバイル社は他の通信事業者から通信回線を借り受け,サービスを提供している MVNO(Mobile Virtual Network Operator)であり,サービスタイプ D では NTT ドコモ社の回線を使用している。月あたり通信量が 3GB を超過すると通信速度が 256kbps に制限されるが,本実験中には速度制限は課されなかった。

クラウドサーバとしては本計測を行うにあたり契約した一台の AWS サーバを用い、大阪大学 敷地内の研究室に設置した Raspberry Pi から ping を用いて応答遅延を計測した。自動的に計測 データを取得できるよう、Raspberry Pi 上で動作する Raspbian において計測用のスクリプトを 用いることで、15 秒毎に時刻を取得し、続けて ping(パケットサイズ 60 バイト、ICMP ECHO メッセージ、パケット数 1)を実行した。計測時刻、ping の出力をログデータとして取り出し、分析を行った。また、日時が応答遅延に与える影響を調べるため、様々な時間帯で計測を行った。計 測は 2/29(土)から 3/27(金)までの四週間に渡って行った。

2 時系列モデルによる回帰

本報告では、時系列モデルとして式 (1) と式 (2) で表される ARMA-GARCH (Autoregressive Moving Average - Generalized Autoregressive Conditional Heteroscedasticity) モデルを用いる.

$$y_t = \sum_{i=1}^p a_i y_{t-i} + \sum_{i=1}^q b_i \varepsilon_{t-i} + c + \varepsilon_t \qquad \varepsilon_t \sim N(0, h_t) \quad i.i.d$$
 (1)

$$h_t = \omega + \sum_{i=1}^r \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^s \beta_i h_{t-i}$$
 (2)

本報告では,各一時間の計測実験において Raspberry Pi が AWS サーバから受け取る ping の応答遅延の実測値を計測時刻順に y_1 , y_2 , ..., y_N と表す.式 (1) において,c は定数項, ε_t はノイズ項であり平均 0 分散 h_t の正規分布に従いそれらは互いに独立であるとし,時刻 t における実測値 y_t は,過去の p 時点前までの実測値の線形和と q 時点前までのノイズ項の加重和と自身のノイズ項によって表される.また,式 (2) において, ω は定数項であり,時刻 t におけるノイズ項が従う正規分布の分散 h_t は,過去の r 時点前までのノイズ項の線形和と s 時点前までのノイズ項が従う正規分布の分散の加重和によって表される.

各パラメータのパラメータ数 (p,q,r,s) は次数と呼ばれ,精度の良い回帰を行うためには対象とする時系列データに応じてこの次数を適切に定める必要がある.しかしながら,計測データのクラスタリングによる分類のためには異なる曜日や時間帯での計測によって得られた時系列データに対して次数を統一して回帰しなければならない.そこで,本報告では,まず,各計測データに対しARMA-GARCH モデルの次数を変えながら適用し,AIC(赤池情報量規準)によってその妥当性を評価することで,各時系列データにとって適切な次数を求める.ここで,AIC は式(3)で定められる指標である.高すぎる次元での回帰による過適合の問題を考慮し,時系列モデルによる回帰のよさを評価する第 1 項とパラメータの少なさを評価する第 2 項を組み合わせている.AIC が小さいほどよいモデルとされている.

$$AIC = -2*(対数尤度) + 2*(パラメータ数)$$
 (3)

次に,各時系列データ間で AIC による最適次数を比較し,その最大のものを用いることとする.次数の探索範囲は $0 \le p \le 2$, $0 \le q \le 2$,r=1, $0 \le s \le 1$ とすると,統一する次数は (p,q,r,s)=(2,2,1,1) となった.

また、応答遅延の変動に注目し、応答遅延の実測値の差分系列 $\{\Delta y_t|y_t-y_{t-1}\}$ に対しても同様にモデルの適用を行った。統一する次数は (p,q,r,s)=(2,2,1,1) となった。

3 AWS サーバと SINET のウェブサーバを対象とした計測デー タのクラスタリング結果の比較

ARMA-GARCH モデルを実測値データもしくは変動値データに対して適用した結果定まるパラメータ W をもとにクラスタリングを行い、時間帯や場所ごとの ping 応答遅延特性を分析する. AWS サーバを対象とした計測データにおける W は、実測値データと差分系列データともに

$$W = [c, a_1, a_2, b_1, b_2, \omega, \alpha_1, \beta_1]$$

表 1: AWS サーバを対象とした実測値の主成分分析における累積寄与率

第一主成分	第二主成分	第三主成分	第四主成分	
0.449	0.681	0.863	0.968	

表 2: AWS サーバを対象とした変動値の主成分分析における累積寄与率

第一主成分	第二主成分	第三主成分	第四主成分	
0.367	0.653	0.794	0.905	

表 3: SINET のウェブサーバを対象とした実測値の主成分分析における累積寄与率

第一主成分	第二主成分	第三主成分	第四主成分
0.324	0.556	0.718	0.834

表 4: SINET のウェブサーバを対象とした変動値の主成分分析における累積寄与率

第一主成分	第二主成分	第三主成分	第四主成分	第五主成分
0.308	0.497	0.648	0.768	0.879

で与えられ、SINET のウェブサーバを対象とした計測データにおける W は、実測値データと差分系列データともに

$$W = [c, a_1, a_2, b_1, b_2, \omega, \alpha_1, \alpha_2, \beta_1, \beta_2]$$

で与えられる。SINET のものに関しては卒論時と合わせています。SINET を対象とした実測値 データにおいて ARMA-GARCH の適用ができないものに関しては省いて行っております。また,AWS サーバを対象とした計測データは 2/29 (土) から 3/27 (金) までに計測したものを用い,SINET のウェブサーバを対象とした計測データは 12/29 (火) から 1/30 (木) までに計測したものを用いる。

ここでは、距離関数をユークリッド距離とし、クラスタ融合手法としてウォード法を用いる. クラスタ数は 6.7.9.12.15 とした.

さらに標準化を行ったパラメータの主成分でのクラスタリングも行う。これは、パラメータ間で大きさや分散が異なること、次元数が多いことによりクラスタリングの精度が低下する可能性を考え、その対処として行った。 AWS サーバと SINET のウェブサーバを対象とした計測データにおける実測値データと変動値データの主成分分析におけるそれぞれの累積寄与率は表 1 から表 4 となっており、累積寄与率が 70% を超えた主成分までを用いてクラスタリングを行う。従って、AWS サーバを対象とした実測値と変動値、SINET のウェブサーバを対象とした実測値に対してはその第三主成分まで、SINET のウェブサーバを対象とした変動値に対してはその第四主成分までを用いる。この時の各主成分の負荷量は表 5 から表 8 となっていた。

以下その結果を図 6 から図 ?? に示す. このそれぞれは, 該当するクラスタリングによって分類された計測データを, 計測した時間帯もしくは曜日ごとに色分けした積み上げ棒グラフ, および計測した曜日と時間帯が同じ計測データを属するクラスタごとに色分けした積み上げ棒グラフである. 前者の縦軸は各クラスタ内の個数に対する割合であり, 上部に各クラスタに属する計測データの個数を示しており, 後者の縦軸は曜日と時間帯が同じデータが属するクラスタの割合であり, 上部に各曜日と時間帯のデータ数を示している.

SINET を対象とした計測データに対する図 6 のクラスタリングにおいて時間帯ごとの図を見ると、オレンジ色や赤色の時間帯が一クラスタ内で占める割合が高くなっているものがあるように見える. しかし、AWS を対象とした計測データに対して同様に実測値を用いてクラスタリングを

表 5: AWS サーバを対象とした実測値の負荷量

第一主成分	第二主成分	第三主成分
0.515	0.133	
-0.418		-0.490
-0.404	-0.166	0.492
0.424		0.475
0.391	0.157	-0.516
-0.119	0.569	0.101
-0.142	0.432	
0.175	-0.643	-0.111
	0.515 -0.418 -0.404 0.424 0.391 -0.119 -0.142	0.515 0.133 -0.418 -0.166 0.424 0.391 0.157 -0.119 0.569 -0.142 0.432

表 7: SINET のウェブサーバを対象とした 実測値の負荷量

	第一主成分	第二主成分	第三主成分
c	0.546		
a_1	-0.374	-0.423	-0.150
a_2	-0.441		0.283
b_1	0.352	0.448	0.154
b_2	0.443	-0.288	
ω		0.349	-0.526
α_1		0.328	-0.294
α_2		0.178	-0.166
β_1	0.163	-0.430	-0.209
β_2			0.708

表 6: AWS サーバを対象とした変動値の負荷量

1.3 ===			
	第一主成分	第二主成分	第三主成分
c			0.721
a_1	0.571		
a_2		-0.190	-0.067
b_1	-0.571		
b_2	0.574		
ω		0.553	-0.127
α_1		0.499	
β_1		-0.629	

表 8: SINET のウェブサーバを対象とした 変動値の負荷量

	第一主成分	第二主成分	第三主成分	第四主成分
c		0.346	0.142	0.465
a_1	-0.552			-0.123
a_2		-0.371	-0.144	-0.482
b_1	0.544	-0.128	-0.103	0.112
b_2	-0.546	0.140		-0.111
ω	-0.204	-0.490	-0.328	0.250
α_1	-0.199	-0.290		0.598
α_2		-0.226	-0.154	-0.248
β_1		0.560	-0.456	-0.132
β_2	0.122	-0.114	0.771	-0.116

行った結果の図 2 ではそのような印象は受けない.ここでは,同様のデータの前処理を行い同一のクラスタリングを行っても,計測対象によって形成されるクラスタに時間ごとの共通の傾向は顕著にみられないようだ.次に曜日ごとでの図を見比べてみたが,SINET か AWS サーバに関わらず,形成されるクラスタに際立った傾向は見られず,どのクラスタにも大体まんべんなく全曜日が含まれている印象だ.最後に時間と曜日ごとにまとめた図に関しては,各時間帯と曜日ごとのデータ数が約 3 程度と少ないが,その中で総数が 4 のものだけに注目してみても,時間と曜日が同じ四つのデータがすべて同じクラスタに属するものはなかった.ただ,三つや二つが同じクラスタに属しているものは見受けられたので,まったく何も傾向がないわけではないように思える.

4 エルボー法による最適クラスタ数の決定

エルボー法を用いて各データセットとそれに対する各クラスタリング手法における最適クラスタ数を定める。各データセットは AWS サーバを対象とした計測で得られた実測値と変動値を ARMA-GARCH (2,2,1,1) で回帰したときのパラメータとその標準化後の主成分の計四つである。クラスタリング手法は非階層的クラスタリングとして k 平均法の距離関数がユークリッド距離, 階層的クラスタリングとしてウォード法、最近傍法、重心法のそれぞれに対して距離関数がユークリッド距

離、キャンベラ距離、マンハッタン距離とした計十個である。要素 $w_i = [w_{i1}, w_{i2}, ..., w_{in}]$ として、

ullet ユークリッド距離: $\sqrt{(oldsymbol{w_i} - oldsymbol{w_j})^2}$

• マンハッタン距離: $\sum_{k=1}^{n} |w_{ik} - w_{jk}|$

• キャンベラ距離: $\sum_{k=1}^{n} \frac{|w_{ik} - w_{jk}|}{|w_{ik}| + |w_{jk}|}$

また,

● 重心法: クラスタ間の重心距離が最小のクラスタ同士を融合する

- 最近傍法: クラスタ間の要素対のうち、最も近い要素間をクラスタ間距離として、クラスタ 間距離が最も小さいクラスタ同士を融合する
- ウォード法:融合後のクラスタ内分散から融合前の二つのクラスタ内分散の差を最小とする という基準でクラスタを融合する

ここでエルボー法とは、クラスタ内誤差分散(SSE: Sum of Squared errors of prediction)と呼ばれるクラスタの集約度をその小ささで表す指標を用いる。クラスタ数を 1 から順に増やしていきながら逐次 SSE を算出し、最適クラスタ数を次のクラスタ数での SSE の変化の割合に対するそのクラスタ数での変化の割合を最大とする箇所と定める。クラスタ数 k での SSE は式 (4) で定められ、最適クラスタ数は式 (5) となる。ここでは、 $2 \le k \le 15$ とした。また、要素 w_i が属するクラスタ重心を $w_G(w_i)$ とし、 $dist(w_i,w_j)$ はそのクラスタリング手法における距離関数上での w_i と w_j の距離である。

$$SSE_k = \sum_{i} dist(\boldsymbol{w_i}, \boldsymbol{w_G}(\boldsymbol{w_i}))^2$$
(4)

最適クラスタ数 =
$$\operatorname*{arg\ max}_{2 \leq k \leq 15} \frac{SSE_k - SSE_{k-1}}{SSE_{k+1} - SSE_k} \tag{5}$$

以下の図は、各データセットと各クラスタリングにおいて、横軸をクラスタ数、縦軸を SSE でとったものを青線の折れ線グラフで示し、赤色の破線の箇所が最適クラスタ数の位置を示している. また、その右図にはクラスタ数を最適クラスタ数としたときのクラスタリングにおける主成分散布図である.

5 クラスタリング手法のクラスタの類似度

作業途中です。あるクラスタリングにおいて同じクラスタに属していた要素のペアが他のクラスタリングにおいてどれだけ同じクラスタに属しているか、また、属していなかったかの割合は計算し終えており、あとはプロットするだけとなっています。

6 スケジュール

日付	予定	
3/31	原稿の考察前までの部分を書き上げる	
4/23	原稿締め切り	
5/21 - 5/22	IN 研究会	

原稿執筆後: 抄録とキーワードの登録(日本語と英語)

プログラム決定後: 参加費支払い 5/14~: 技報ダウンロード開始

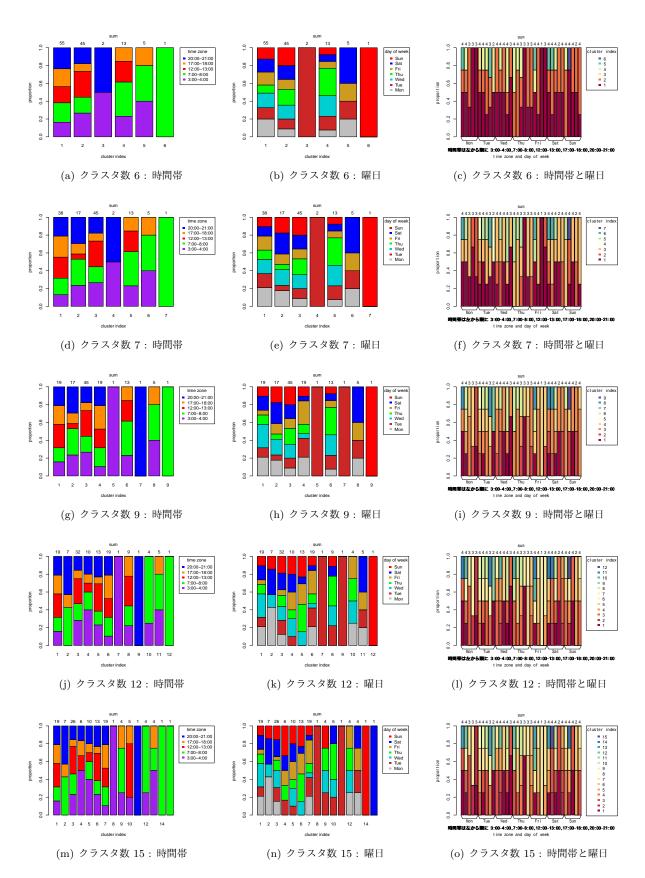


図 2: AWS サーバを対象とした計測の実測値によるクラスタリング

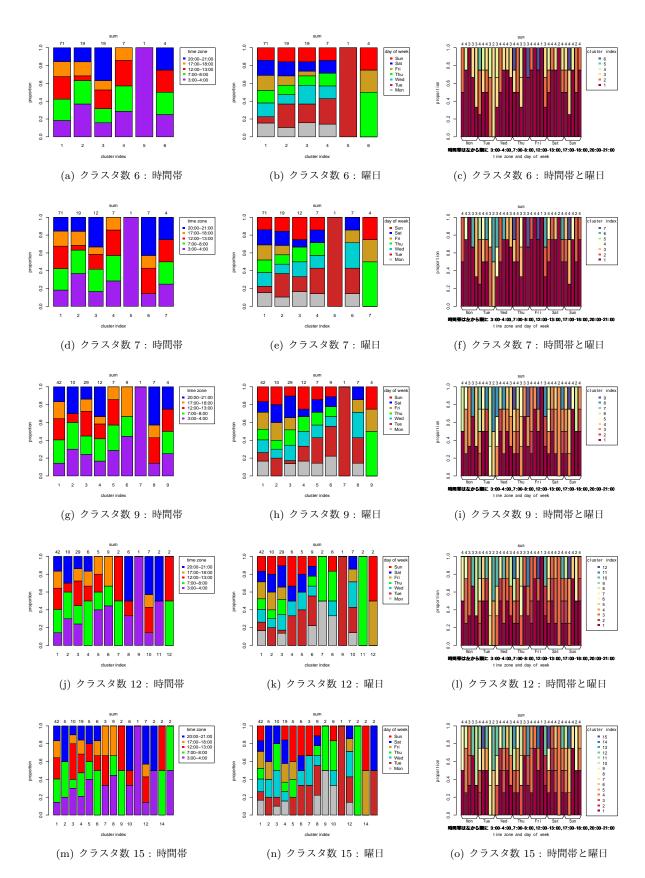


図 3: AWS サーバを対象とした計測の変動値によるクラスタリング

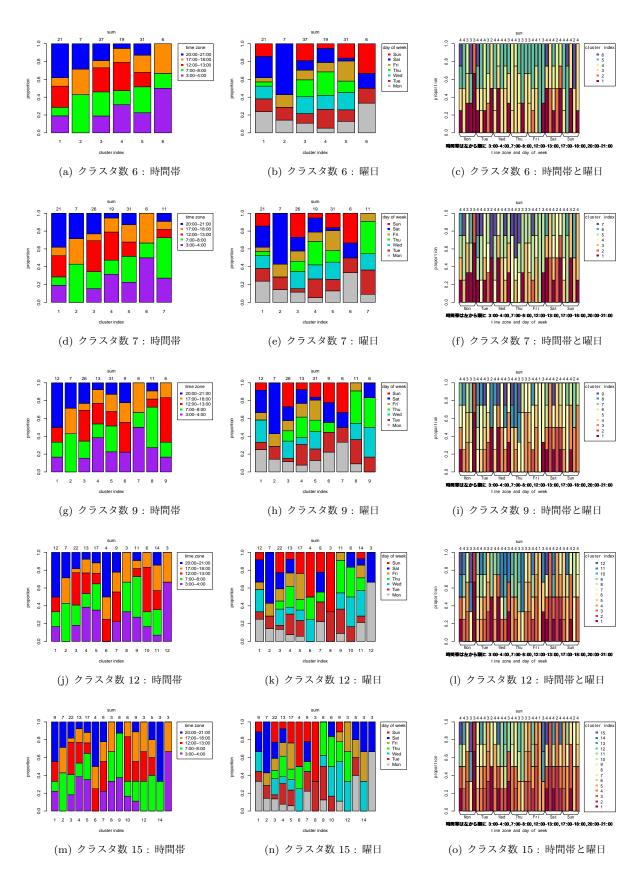


図 4: AWS サーバを対象とした計測の実測値の主成分によるクラスタリング

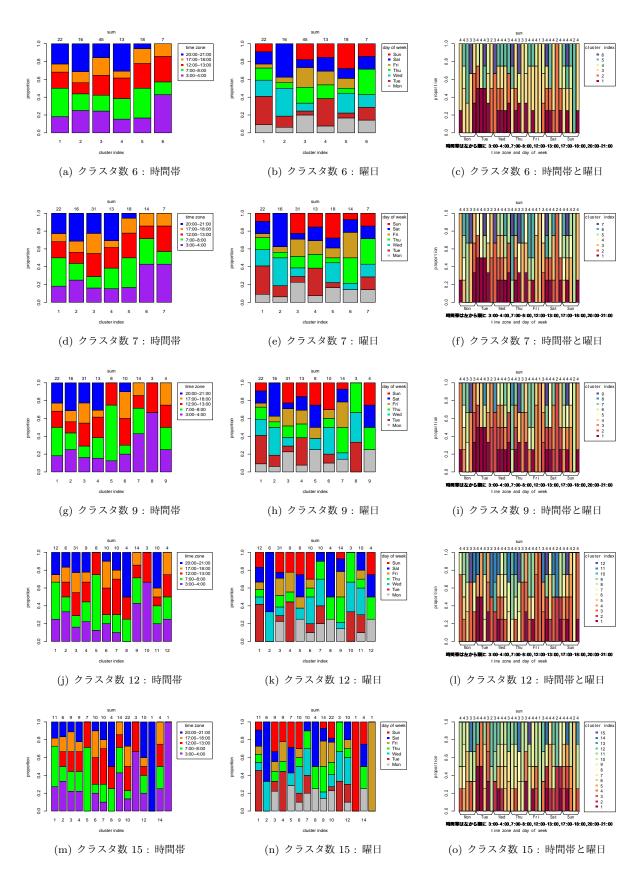


図 5: AWS サーバを対象とした計測の変動値の主成分によるクラスタリング

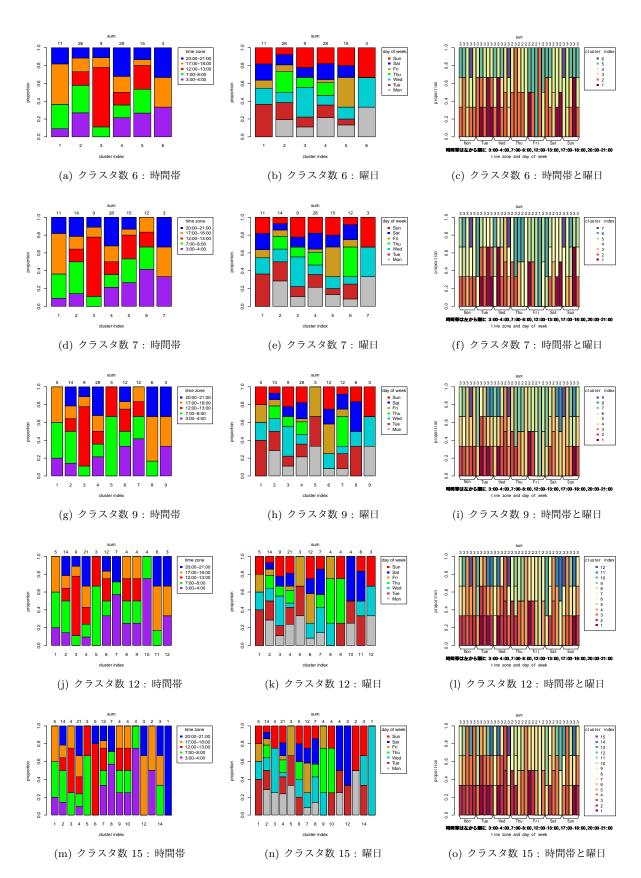


図 6: SINET のウェブサーバを対象とした計測の実測値によるクラスタリング

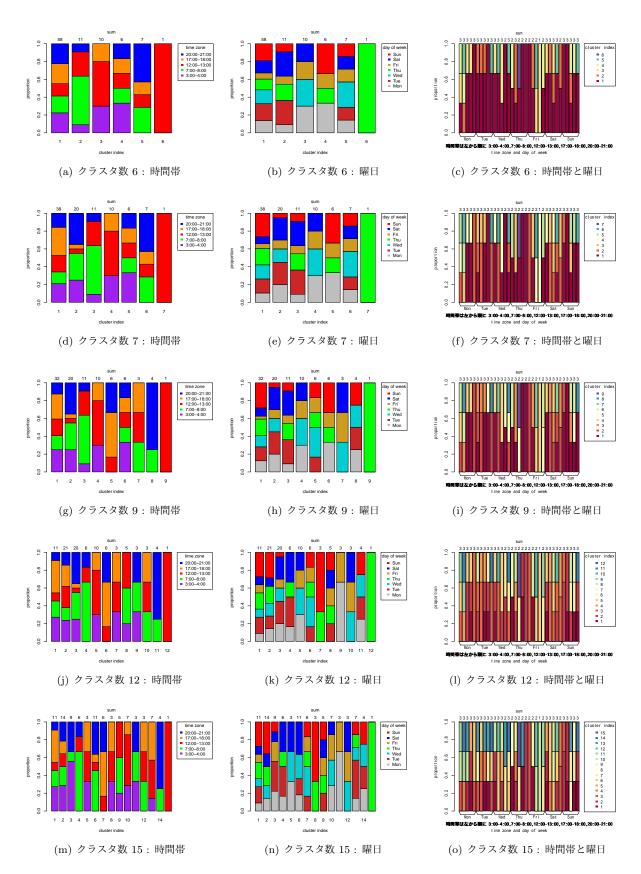


図 7: SINET のウェブサーバを対象とした計測の変動値によるクラスタリング

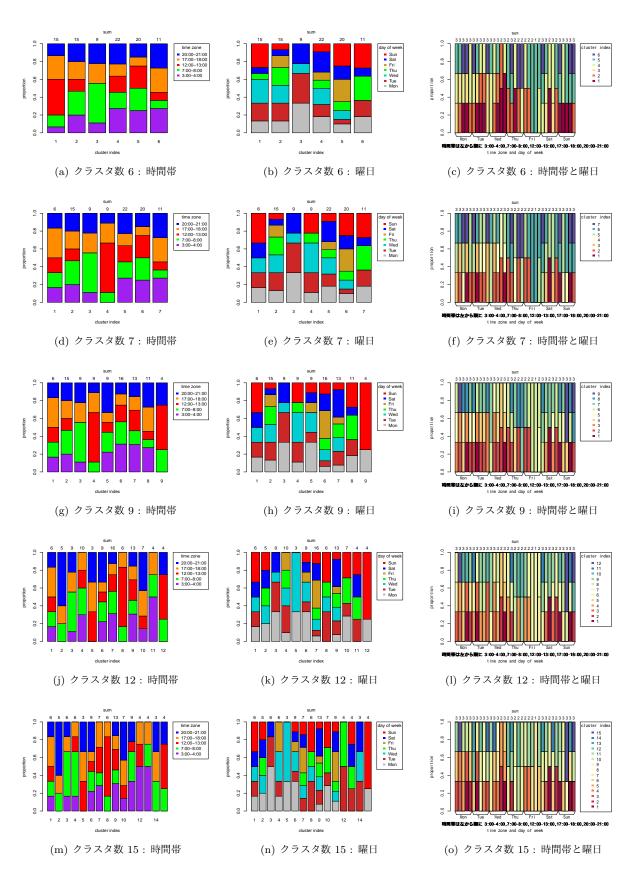


図 8: SINET のウェブサーバを対象とした計測の実測値の主成分によるクラスタリング

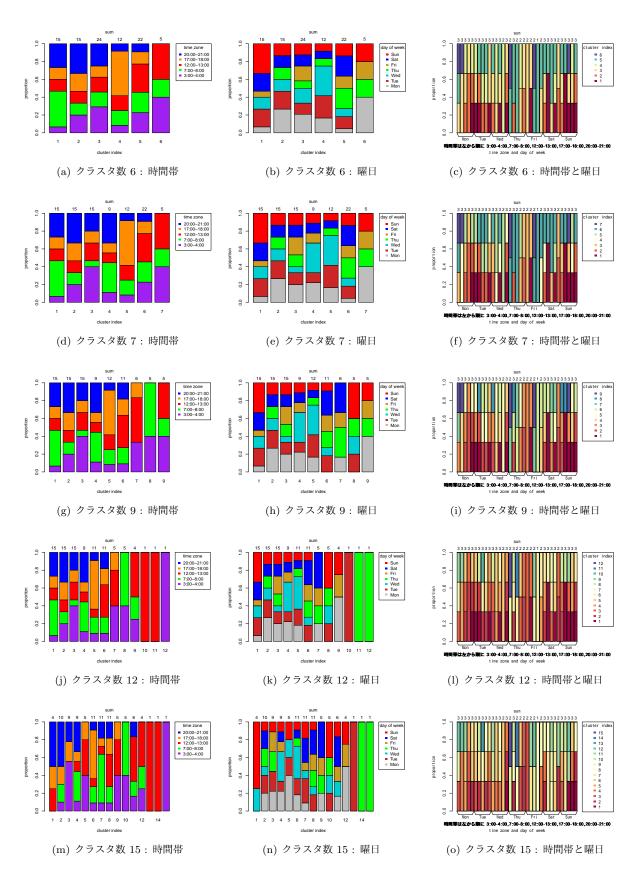


図 9: SINET のウェブサーバを対象とした計測の変動値の主成分によるクラスタリング

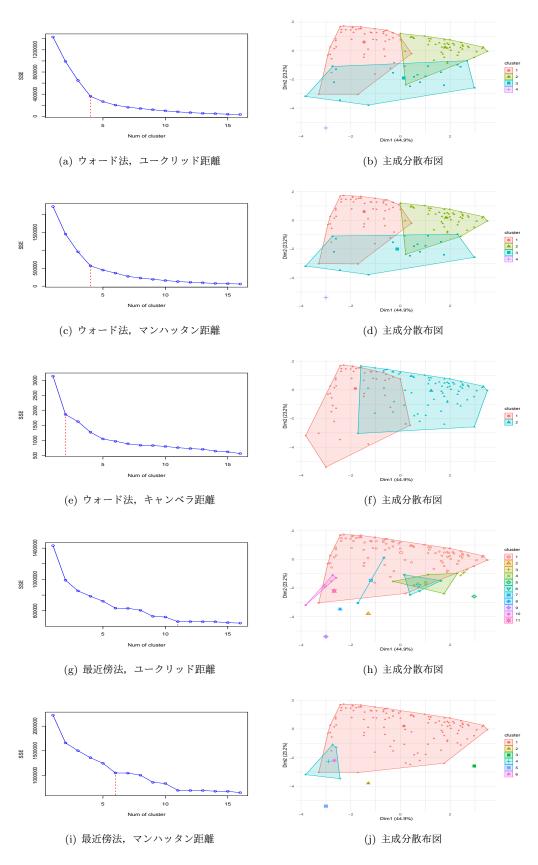


図 10: 実測値によるクラスタリング

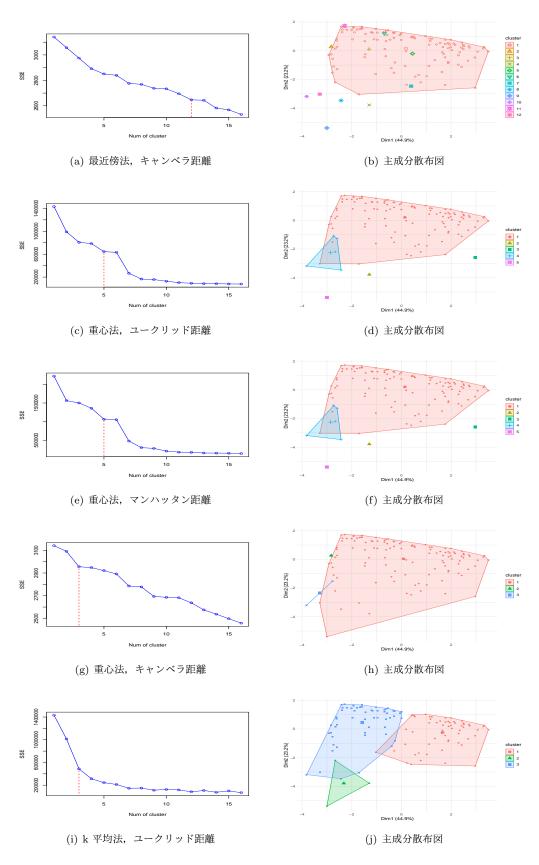


図 11: 実測値によるクラスタリング

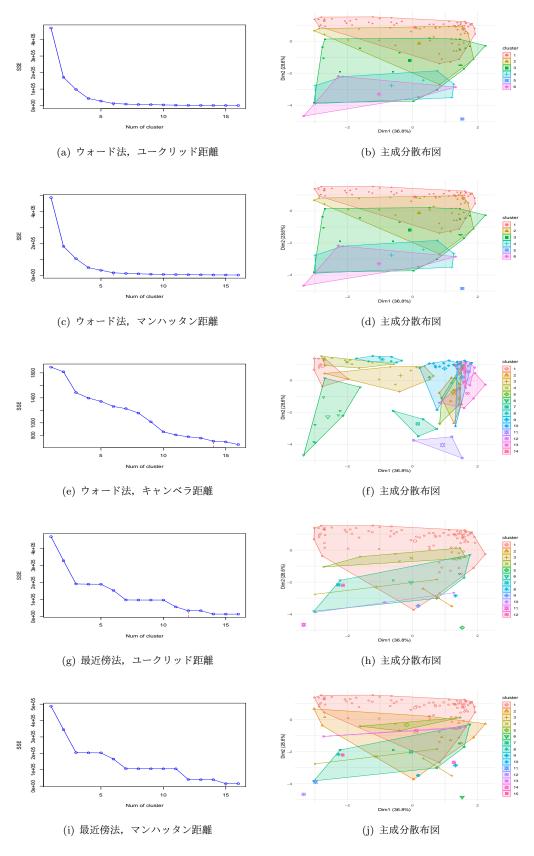


図 12: 変動値によるクラスタリング

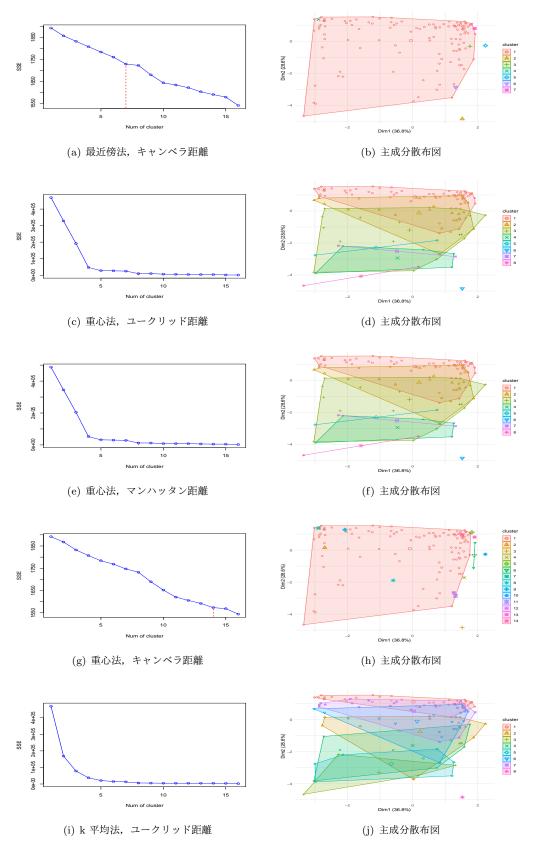


図 13: 変動値によるクラスタリング

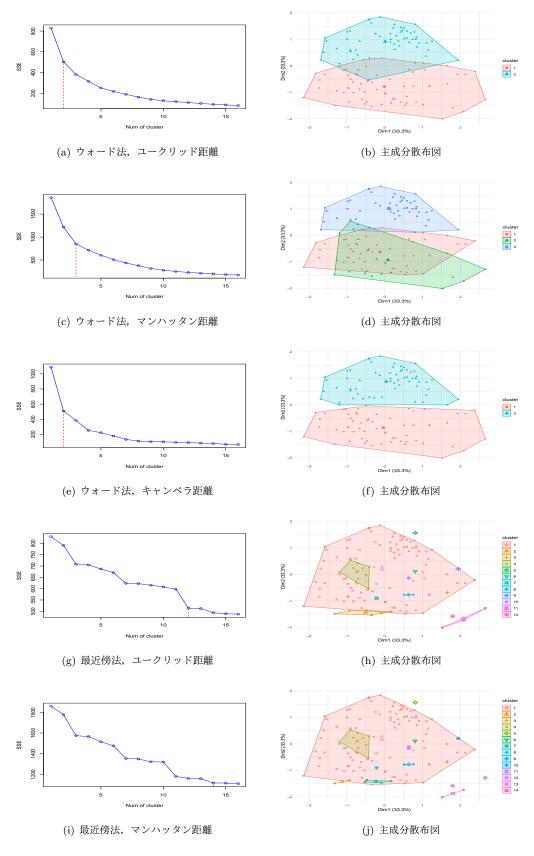


図 14: 実測値の主成分によるクラスタリング

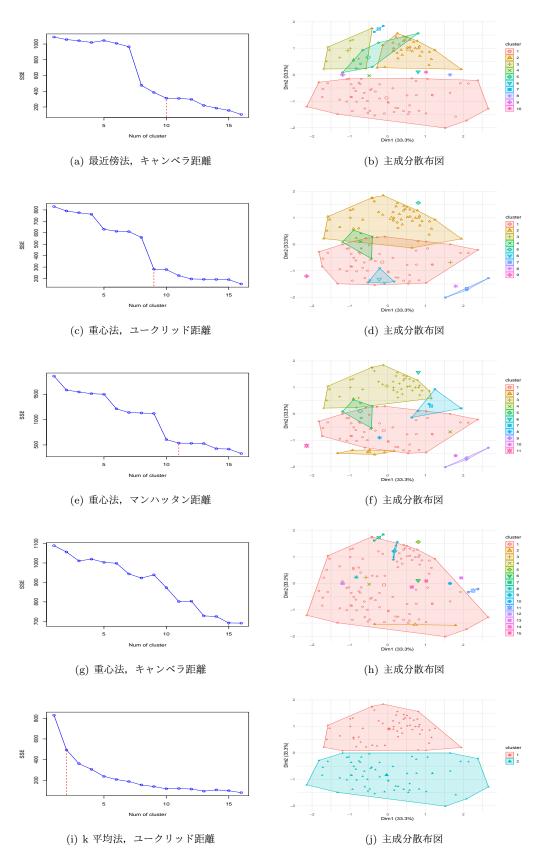


図 15: 実測値の主成分によるクラスタリング

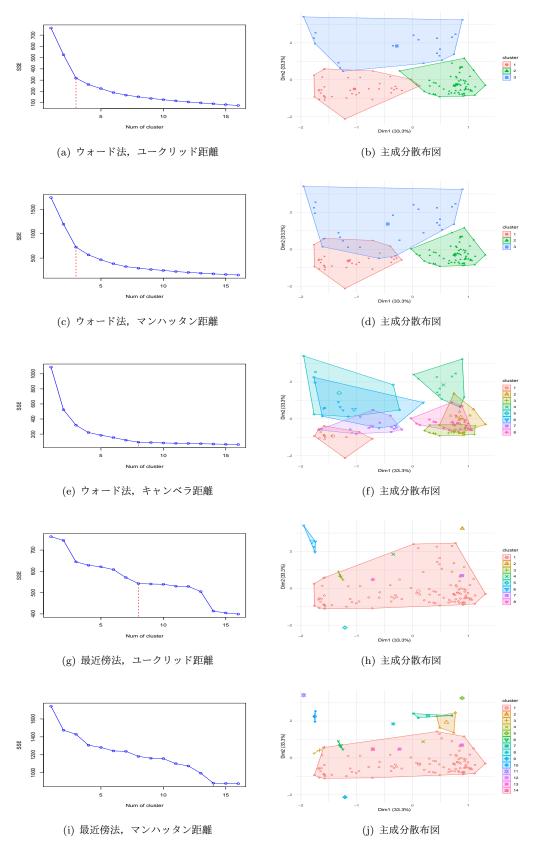


図 16: 変動値の主成分によるクラスタリング

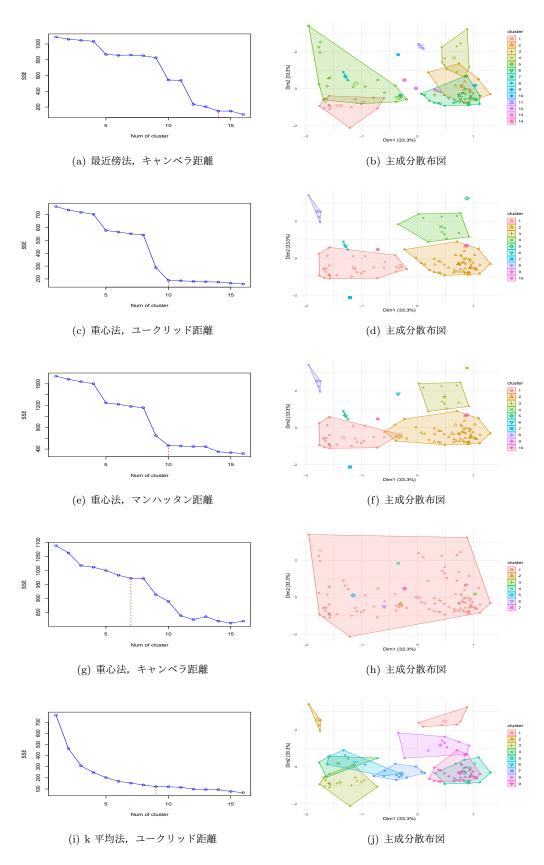


図 17: 変動値の主成分によるクラスタリング