

# LTE 環境における応答遅延特性の時系列モデリングによる分析

山本 航平<sup>†</sup> 若宮 直紀<sup>†</sup> 中野 亮<sup>††</sup> 藤原 亮介<sup>††</sup>

<sup>†</sup> 大阪大学大学院情報科学研究科バイオ情報工学講座

〒565-0871 大阪府吹田市山田丘 1-5

<sup>††</sup> 株式会社日立製作所 研究開発グループ

〒185-8601 東京都国分寺市東恋ヶ窪 1-280

E-mail: <sup>†</sup>{k-yamamt,wakamiya}@ist.osaka-u.ac.jp, <sup>††</sup>{ryo.nakano.xd,ryosuke.fujiwara.mb}@hitachi.com

あらまし 産業用モニタリングシステムにおける運用管理コストの低減のため、迅速な障害検知・予測、ならびに原因の特定と対処法の提示が求められている。その実現にむけて、無線機器からサーバまでの LTE 回線を含む通信路について、異なる時間帯において応答遅延の計測を行った。本稿では、遅延の変動特性や、曜日や時間帯に依存した傾向の有無について、時系列モデルを用いたクラスタリングによって分析した結果を示す。

キーワード Long Term Evolution, 応答遅延, 時系列モデリング, 異常検知

## Analysis of delay characteristics of connection through LTE by time series analysis

Kohei YAMAMOTO<sup>†</sup>, Naoki WAKAMIYA<sup>†</sup>, Ryo NAKANO<sup>††</sup>, and Ryosuke FUJIWARA<sup>††</sup>

<sup>†</sup> Department of Bioinformatic engineering, Graduate school of Information Science and Technology, Osaka University

1-5 Yamadaoka, suite-shi, Osaka, 565-0871, Japan

<sup>††</sup> Research & Development Group, Hitachi, Ltd.

1-280 Higashikoigakubo, kokubunji-shi, Tokyo, 185-8601, Japan

E-mail: <sup>†</sup>{k-yamamt,wakamiya}@ist.osaka-u.ac.jp, <sup>††</sup>{ryo.nakano.xd,ryosuke.fujiwara.mb}@hitachi.com

**Abstract** In order to reduce the costs of operation and management in industrial monitoring systems, it is necessary to detect or predict failures quickly, identify the cause, and present countermeasures. To achieve this, we measured response delays of connection from one wireless device to one server, including LTE, in different time zones. In this technical report, we show the results of analyzing the fluctuation characteristics of response delay and the tendency depending on the day of week or time zone by clustering using a time series analysis.

**Key words** Long Term Evolution, response delay, time series analysis, anomaly detection

### 1. はじめに

近年, IoT (Internet of Things) 技術の発展とともに産業用モニタリングシステム [1] が普及している。これは、工場などの産業施設に設置された機器から直接、あるいは配置したカメラやセンサーなどの IoT デバイスを通じて、機器の稼働状態に関する情報をリアルタイムで収集し、キャリア回線を通じてクラウドサーバに送信、データ処理を行い、運用管理担当者に提示するものである。作業員の巡回による工場内の機器の点検業務を自動化できるため、人員コストの軽減、目視確認より生じる人的ミスの低減、リアルタイムなデータの活用などの効

果が期待できる。一方で、長期間の運用のなかで機器に具備された IoT デバイスの故障、工場内ネットワークの通信の途絶、クラウドサーバへの通信の遅延などの障害は避けられない。このような障害が発生した場合には工場内の機器の稼働状況を把握することが困難となり、工場の稼働停止や業務の遅れなどが引き起こされ大きな損失をもたらす可能性がある。

障害発生時には迅速な復旧作業が求められるが、障害はその原因や内容、規模によってさまざまである。しかしながら、現状ではシステムから得られる情報を用いてこれらの障害を適切かつ迅速に区別することが困難であるため、障害発生時には運用管理担当者が直接現場に行き障害の原因や内容、規模を確認す

る必要があり、多大な運用管理コストが発生している。そのため、モニタリングシステムによる運用管理コストの低減のためには、直接また間接的に取得できる情報にもとづいて、障害発生を迅速に検知また予測するとともに、障害の原因を特定し、加えてその対処法を示すことが求められている。

すでに我々の研究グループでは機器間で測定される受信電波強度の時間変化にもとづいて空間的な電波伝搬特性の変動を推定することにより工場内での無線通信の異常を検知する手法を提案している [2]。そこで、本稿では工場内で機器稼働情報を収集する無線機器からクラウドサーバまでの通信路で発生する異常の検知手法の確立を目的とし、その第一段階としてさまざまな曜日、時間帯で応答遅延を計測し、遅延変動の特性の分析を行う。キャリア回線としては産業用モニタリングシステムに広く用いられている LTE (Long Term Evolution) 回線を用いる。LTE 回線の特性評価に関する研究としては、応答遅延の他の無線回線との比較評価 [3] [4] や TCP パケットの振る舞いに関する調査 [5]、人の混雑状況と遅延分布に関する分析 [6] などが行われている。本稿では、産業用モニタリングシステムを想定した 4 週間にわたる継続的な計測実験を行い、得られた応答遅延を時系列モデリングやクラスタリングによって分析し、その特性を見いだす。

第 2 章では本稿で実施した計測実験の設定について述べる。第 3 章では時系列モデルによる回帰について述べる。第 4 章では回帰結果に用いたクラスタリングにもとづく分析について述べる。第 5 章では全体のまとめと今後の課題について述べる。

## 2. 計測実験の設定

本稿では図 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実験やシステム開発のために契約した一台の 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、3 時, 7 時, 12 時, 17 時, 20 時のそれぞれ 1 時間において計測を行った。それぞれの時間帯ごとに得られた計測値を区間データと呼ぶ。計測は 2/29 (土) から 3/27 (金) までの四週間に渡って行った。したがって、区間数は 140、総計測数は 33600

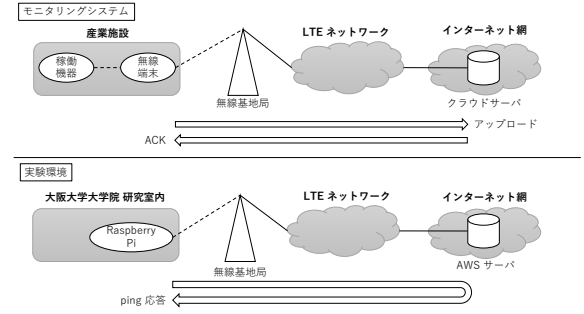


図 1 モニタリングシステムと実験環境の対応図

となるが、Raspberry Pi の動作不良等により計測データが欠損した区間を除く 122 区間の計 29280 の計測値について分析を行った。

Raspberry Pi の動作不良のため未使用の計測データ

3/6 (金) 17:00-18:00, 3/7 (土) 17:00-18:00  
 3/12 (木) 12:00-13:00, 3/12 (木) 17:00-18:00  
 3/13 (金) 17:00-18:00, 3/14 (土) 17:00-18:00  
 3/15 (日) 17:00-18:00, 3/16 (月) 17:00-18:00  
 3/17 (火) 20:00-21:00, 3/20 (金) 17:00-18:00  
 3/22 (日) 17:00-18:00, 3/23 (月) 12:00-13:00  
 3/23 (月) 20:00-21:00, 3/24 (火) 17:00-18:00  
 3/24 (火) 20:00-21:00, 3/25 (水) 20:00-21:00  
 3/26 (木) 20:00-21:00, 3/27 (金) 20:00-21:00

## 3. 時系列モデルによる回帰

本稿では、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 [7] を用いる。

$$y_t = \sum_{i=1}^p a_i y_{t-i} + \sum_{i=1}^q b_i \varepsilon_{t-i} + c + \varepsilon_t \quad (1)$$

$$\varepsilon_t \sim N(0, h_t) \quad \text{i.i.d}$$

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

時系列モデルによる回帰では、適切な次数  $p, q, r, s$  のもとで実測値  $y_t$  の時系列を最も精度良くモデル化できるパラメータ  $a_i, b_i, c, \omega, \alpha_i$ , および  $\beta_i$  を算出する。ここで、 $y_t$  ( $1 \leq t \leq N, N = 240$ ) は 1 時間の計測区間のそれぞれにおける計測時刻順の実測値である。また、 $c$  は定数項、 $\varepsilon_t$  は平均 0、分散  $h_t$  の独立同一な正規分布に従うノイズ項である。したがって、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と  $q$  時点前までのノイズ項のそれぞれの重み付き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、 $\omega$  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と  $s$  時点前までのノイズ項が従う正規分布の分散のそれぞれの重み付き和によって表

表 1 最適次数と統一次数での対数尤度の比較

	最適次数での対数尤度	統一次数での対数尤度
実測値データ	-969.8327	-971.9196
変動値データ	-924.6495	-922.7543

される。

次数  $(p, q, r, s)$  は対象とする時系列データに応じて適切に定める必要がある。最適な次数は区間データごとに異なるが、クラスタリングによる分類、分析のために共通の次数を用いることとする。次数  $0 \leq p \leq 2, 0 \leq q \leq 2, r = 1$ , および  $0 \leq s \leq 1$  のそれぞれの組み合わせに対して AIC (赤池情報量規準) [8] [9] を求めたところ、最大次数である  $(p, q, r, s) = (2, 2, 1, 1)$  が最適な計測区間が存在することから、これを共通の次数として用いることとした。なお、式 (1) における  $p, q$  は次元数を抑えるために 2 まで、また、式 (2) における  $r, s$  は一般的に十分な性能が得られる 1 までとした [10]。ただし、 $r = 0$  は回帰に用いた R 言語のパッケージ fGarch で設定できないため除外した。また、実測値  $y_t$  の代わりに実測値の差分である変動値の系列  $\{\Delta y_t | y_t - y_{t-1}\}$  による時系列解析についても検討を行ったところ、同様に  $(p, q, r, s) = (2, 2, 1, 1)$  を用いることとなった。なお、AIC による最適次数よりも大きい次数を適用することの回帰精度への影響を検証するため、実測値における最適次数が  $(0, 1, 1, 0)$  である区間データに対して次数  $(2, 2, 1, 1)$  を適用した際の対数尤度の比較したところ、その変化量は約 0.2% であった (表 1)。

図 2 に月曜日の 7:00~8:00 の時間帯において得られた実測値に対する回帰結果を示し、表 2 にこれらの時系列データに対するモデルのパラメータを示す。図 2 より、回帰線は実測値のベースラインを表現していることが読み取れる。この回帰線の変動の仕方が表 2 で定められる。表より、 $a_1$  と  $a_2$  はともに負の値をとっていることから、一時点前および二時点前の実測値が大きいほど推定値は小さくなることがわかる。これは一時的に大きな実測値が発生しても実測値は増加傾向にはならず、ベースラインに回帰されることを意味する。 $a_1, a_2$  ともに (a) よりも (b) の方が小さいことから、(b) の推定値は (a) のものよりも過去の実測値との間の負の相関が強い傾向を示した。また、 $b_1$  と  $b_2$  はともに正の値をとっていることから、一時点前および二時点前のノイズ項が大きいほど推定値が大きくなることがわかる。これは、過去のノイズの推定値に対する影響が時間的に持続することを意味する。 $b_1, b_2$  ともに (a) よりも (b) の方が大きいことから、(b) のノイズの影響は (a) のものよりも残りやすい傾向を示した。さらに、 $a_1$  と  $a_2$  の絶対値は  $a_2$  の方が大きく、 $b_1$  と  $b_2$  の絶対値も  $b_2$  の方が大きいので、推定値には二時点前の実測値やノイズが一時点前のものよりも相関が強いと考えられる。また、定数項  $c$  は推定値の基準値と考えられるため、(b) の方が (a) よりも推定値が大きくなりやすい傾向があると読み取れる。さらに、 $a_1, a_2, b_1, b_2, c$  のそれぞれの絶対値の大きさは (b) が (a) よりも大きいことが読み取れる。このことは、(b) の推定値は (a) よりも変動が激しいこ

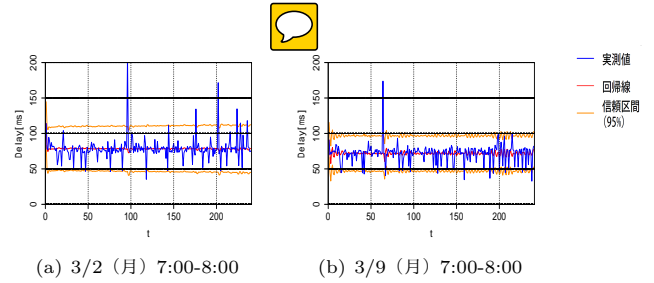


図 2 月曜日 7:00-8:00 の実測値に対する回帰線

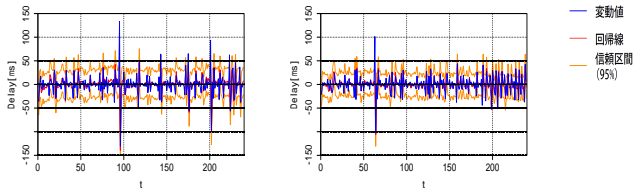
表 2 図 2 に対するモデルのパラメータ

	(a) 3/2 (月) 7:00-8:00	(b) 3/9 (月) 7:00-8:00
$a_1$	-0.06019205	-0.19235040
$a_2$	-0.11082113	-0.76329719
$b_1$	0.01858739	0.17460199
$b_2$	0.06935663	0.65314854
$c$	91.83487587	140.40728113
$\omega$	0.20250175	8.66125582
$\alpha_1$	0.00000001	0.00000001
$\beta_1$	0.99999999	0.94672654

とを意味する。また、ノイズ項が従う分散の基準点  $\omega$  は (b) の方が (a) よりも大きいため、(b) の方が (a) よりもノイズの分散が大きくなりやすいことを意味する。一方で、 $\alpha_1, \beta_1$  は (a) と (b) との間に大きな差はなかった。つまり、一時点前のノイズ項が従う分散やノイズの大きさが推定されるノイズ項の分散へ与える影響は (a) と (b) で大差ないことを意味する。さらに、 $\alpha_1$  の値はかなり小さいため、一時点前のノイズが推定されるノイズ項の分散に与える影響は限りなくゼロに等しいと読み取れる。

またここで、図 2 から、単発的に大きな応答遅延が発生することが読み取れるが、過去の計測データを用いて回帰を行う時系列モデルではこのようなスパイクの発生を予測することが困難であることが読み取れる。

次に図 3 に同区間の変動値に対する回帰結果を示し、表 3 にそのモデルのパラメータを示す。図 3 より、実測値に対する回帰と同様に、変動値に対する回帰は変動値のベースラインを表現していることが読み取れる。表より、 $a_1$  と  $a_2$  はともに負の値をとっていることから、推定値は一時点前および二時点前の変動値と負の相関を持つことがわかる。一方で、(a) と (b) ともに  $a_2$  よりも  $a_1$  の方が小さいことから、推定値は一時点前の変動値が正なら負に、負なら正になりやすい傾向を示した。この点において、変動値は正負に激しく動くものの長期的な変動は 0 であると考えられる。また、 $b_1$  と  $b_2$  はともに負の値をとっていることから、一時点前および二時点前のノイズが大きいほど推定値は小さくなることがわかる。しかし、(a) においては  $b_2$  よりも  $b_1$  の方が小さく、(b) においては  $b_1$  よりも  $b_2$  の方が小さかったことから、(a) と (b) の過去のノイズによる変動の仕方に違いが見られた。c に関しては (a) と (b) に大差は見られない。ノイズが従う分散の回帰に関わるパラメータ  $\omega$  は (a) の方が (b) よりも大きいことから、(a) のノイズの分散は大きくなりやすいことを意味する。 $\alpha_1$  は実測値同様に (a)



(a) 3/2 (月) 7:00-8:00 (b) 3/9 (月) 7:00-8:00

図 3 月曜日 7:00-8:00 の変動値に対する回帰線

表 3 図 3 に対するモデルのパラメータ

	(a) 3/2 (月) 7:00-8:00	(b) 3/9 (月) 7:00-8:00
$a_1$	-0.22605053	-0.67683663
$a_2$	-0.06377989	-0.01662159
$b_1$	-0.90702979	-0.42991003
$b_2$	-0.17145273	-0.67824234
$c$	-0.07518171	-0.08615516
$\omega$	166.42677036	8.33052597
$\alpha_1$	0.00000001	0.00000001
$\beta_1$	0.25766368	0.94593359

と (b) との間に大きな差はなく、その値は限りなくゼロに近い  
ため、一時点前のノイズの大きさが推定されるノイズ項の分散  
に与える影響はほとんどなかった。また、 $\beta_1$  は (a) よりも (b)  
の方が大きいことから、(b) の方が (a) よりも一時点前のノイズ  
が従う分散の大きいほど推定されるノイズの分散が大きくなる  
ことがわかる。

以上の結果より、時系列モデルを用いることにより、実測値  
並びに変動値のベースラインの変動の仕方をパラメータを用い  
て定量化することができる。したがって、時系列モデリングに  
よる異常検知においては、前もって定めたモデルパラメータの  
もとで回帰予測を行っておき、そのベースラインの予測値と実  
際の計測値を対比させ、結果が大きく外れ続けた場合に異常を  
検知するものや、工場内部の無線端末上で一定時間幅ごとにリ  
アルタイムに回帰分析を行い、その結果定まるモデルパラメー  
タをベースラインの変化の時間依存性を定量化したものとみな  
し、前の時間幅におけるモデルパラメータと大きく異なった場  
合に異常を検知するものなどが考えられる。

#### 4. クラスタリングにもとづく分析

ARMA-GARCH モデルを実測値または変動値の区間データ  
に適用して得られるパラメータ  $\mathbf{W} = [a_1, a_2, b_1, b_2, c, \omega, \alpha_1, \beta_1]$   
をもとにクラスタリングを行い、曜日や時間帯の異なる計測  
結果の類似性や傾向を分析する。まず、各パラメータの分布が  
異なるため、平均 0、標準偏差 1 となるように標準化を行う。  
具体的には、パラメータ  $\mathbf{W} = [w_1, w_2, \dots, w_8]$  について、区間  
データ  $j$  のパラメータを  $\mathbf{W}_j = [w_{j1}, \dots, w_{j8}]$ 、パラメータ  $w_{ji}$   
の区間データ間の平均を  $\mu_i$ 、標準偏差を  $\sigma_i$  とすると、区間  
データ  $j$  の標準化後のパラメータ  $\mathbf{W}'_j = [w'_{j1}, \dots, w'_{j8}]$  は次式で  
得られる。

$$w'_{ji} = \frac{w_{ji} - \mu_i}{\sigma_i}$$

以降では簡単のため、このようにして得た標準化後のパラメー

表 4 累積寄与率

(a) 実測値パラメータに対して

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

(b) 変動値パラメータに対して

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 5 主成分負荷量

(a) 実測値パラメータに対して

	第一主成分	第二主成分	第三主成分
$c$	0.515	0.133	
$a_1$	-0.418		-0.490
$a_2$	-0.404	-0.166	0.492
$b_1$	0.424		0.475
$b_2$	0.391	0.157	-0.516
$\omega$	-0.119	0.569	0.101
$\alpha_1$	-0.142	0.432	
$\beta_1$	0.175	-0.643	-0.111

(b) 変動値パラメータに対して

	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		
$\omega$		0.553	-0.127
$\alpha_1$		0.499	
$\beta_1$		-0.629	

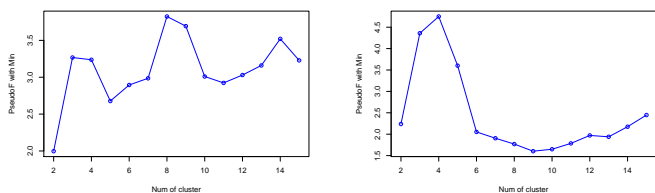
タをそれぞれ実測値パラメータ、変動値パラメータと呼ぶ。

次に、実測値パラメータ、変動値パラメータを主成分分析 [11]  
することで次元を削減する。これは、実測値パラメータ、変動  
値パラメータの次元が高いことによって、一部のパラメータの  
細かな差異が重視されて本来同一の傾向がある区間データが異  
なるクラスタに収容されるなどの問題を抑制するためである。  
実測値、変動値それぞれのパラメータ  $\mathbf{W}'$  に対して主成分分析  
を行った結果の累積寄与率を表 4 に示す。実測値パラメータ、  
変動値パラメータのいずれにおいてもおよそ 80% の累積寄与  
率が得られている第三主成分までを用いてクラスタリングを行  
うこととする。また、各主成分の主成分負荷量を表 5 に示す。  
表より、パラメータ  $a_1, b_1, b_2$  が実測値、変動値のいずれに  
対しても第一主成分に大きく寄与していることがわかる。従っ  
て、実測値、変動値の時間変化においては、一時点前の実測値  
や変動値、また、二時点前までのノイズ項が関係していると思  
えられる。実測値パラメータについてはさらに定数項  $c$  と二時  
点前の実測値が影響している。

クラスタリングには、距離関数としてユークリッド距離を、  
また、階層クラスタリング手法の一つであるワード法 [12] を  
用いる。なお、距離関数としてはキャンベラ距離、マンハッ  
タン距離、クラスタリング手法としては k-means 法、最近隣法、  
重心法についても試したが上記の組み合わせが最もデータの分  
離性能が高かった。ワード法では、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手法である。クラスタ内分散は、クラ  
スタ重心を基準としたクラスタの要素のばらつきとして算出さ  
れる。

クラスタリングによってデータを分析するためには適切なク  
ラスタ数を定める必要がある。本研究においては、クラスタリ





(a) 実測値パラメータの主成分に対して (b) 変動値パラメータの主成分に対して

図 4 各クラスタ数における Pseudo F with Min

ング指標の一つである Pseudo F (Calinski Harabasz 基準) [13] に我々の研究グループで改良を加えた Pseudo F with Min [14] を用いて定量的に最適なクラスタ数を決定する. Pseudo F はクラスタ間分散のクラスタ内分散に対する比として与えられ, クラスタ間の離散性が高く, クラスタ内の凝集性が高い場合に値が大きくなる. Pseudo F におけるクラスタ間分散は, 全要素の代表点を基準としたクラスタの代表点のばらつきで定義されており, 形成されるクラスタ間の距離を反映していない. そのため, Pseudo F の評価値が高くても, クラスタ間が密である可能性がある. そこで, クラスタ間分散ではなく最近傍クラスタとの距離を用いるのが Pseudo F with Min であり, 式 (3) で評価値が得られる.

$$\frac{\sum_{i=1}^k n_i \min\{dist(m_i, m_j)^2, j \neq i\}}{1 + \sum_{i=1}^k \sum_{x \in C_i - \{m_i\}} dist(x, m_i)^2} \quad (3)$$

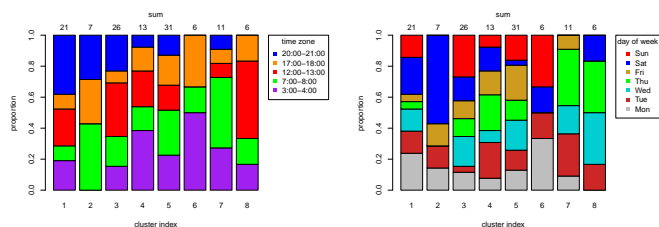
ここで,  $k$  はクラスタ数,  $C_1, \dots, C_k$  はクラスタ集合を表し, 要素数  $n_i = |C_i|$ , 全要素数  $N = \sum_{i=1}^k n_i$  である. また,  $m_i$  はクラスタ  $i$  の代表点であるメドイド [15],  $m$  は全要素に対するメドイドである. 要素集合  $C$  のメドイドは式 (4) で与えられる.

$$\arg \min_{x \in C} \sum_{y \in C - \{x\}} dist(x, y) \quad (4)$$

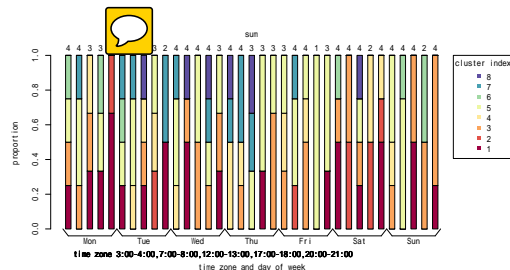
実測値パラメータおよび変動値パラメータの主成分についてクラスタ数を変化させて Pseudo F with Min を求めた結果を図 4 に示す. 図より, 実測値パラメータの場合にはクラスタ数 8 で, 変動値パラメータの場合にはクラスタ数 4 でそれぞれ Pseudo F with Min が最大になることがわかる.

実測値および変動値のそれぞれにおいて, クラスタ数に対する Pseudo F with Min の値は図 4 となった. 図 4(a) より, 実測値における Pseudo F with Min の値はクラスタ数 8 で最大となっており, 図 4(b) より, 変動値におけるものはクラスタ数 4 で最大となることが読み取れる. したがって, 実測値の主成分を用いたクラスタリングにおけるクラスタ数は 8, 変動値の主成分を用いたクラスタリングにおけるクラスタ数は 4 とする.

このもとでクラスタリングを行った結果を図 5 と図 6 に示す. 図 (a) は, 横軸に示すそれぞれのクラスタについて, 各時間帯の区間データが占める割合を, また図 (b) では, 各曜日の区間データが占める割合を, それぞれ積み上げグラフで示している. なお, クラスタごとの区間データ数はそれぞれの上部に

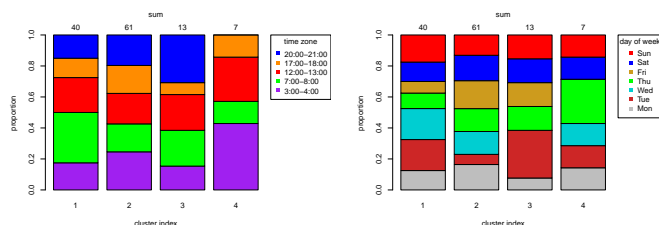


(a) 時間帯ごとに属するクラスタ (b) 曜日ごとに属するクラスタ

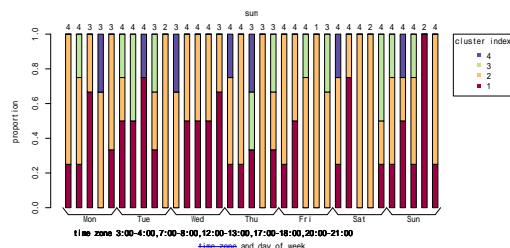


(c) 各時間帯と曜日に属するクラスタ

図 5 実測値パラメータの主成分をもとにクラスタ数 8 で行ったクラスタリング結果



(a) 時間帯ごとに属するクラスタ (b) 曜日ごとに属するクラスタ



(c) 各時間帯と曜日に属するクラスタ

図 6 変動値パラメータの主成分をもとにクラスタ数 4 で行ったクラスタリング結果

示すとおりであり, 例えば実測値パラメータの主成分によるクラスタリングでは 21, 7, 4 であった. また, 図 (c) では, 横軸を曜日ごと, さらに時間帯で区切り, 各クラスタに属する区間データの割合を積み上げグラフで示している. 各曜日, 時間帯の区間データ数を上部に示す.

図 5(a) より, いずれのクラスタにおいてもすべての時間帯の区間データが一樣に含まれているということではなく, 属する区間データには時間帯ごとの偏りがあることがわかる. 特に, 図 5(c) より全ての曜日ににおいて 20:00-21:00 の区間データのうち少なくとも一つはクラスタ 1 に属していることから, 20:00-21:00 の区間データにはなんらかの共通性があると考えられる. さらにクラスタ 1 とクラスタ 3 には, 昼休憩時や帰

宅時間帯、一日の休憩時といった利用者が多いと考えられる 12:00 - 13:00, 17:00-18:00, 20:00-21:00 の計測データの割合が多い。さらに、休日の利用者が多いと考えられる土曜日と日曜日の割合が多い。したがって、クラスタ 1 と 3 には、利用者が多い状況で計測されたデータが属しやすい傾向があるのではないかと考えられる。一方クラスタ 7 は、深夜帯や一日の起床時間帯といった利用者が少ないと考えられる 3:00 - 4:00, 7:00 - 8:00 の計測データが多く、クラスタ 7 には休日による利用者が多いと考えられる土曜日と日曜日に計測されたデータは属していなかった。したがって、クラスタ 7 には、利用者が多くない状況で計測されたデータが属しやすい傾向があるのではないかと考えられる。クラスタ 2, 6 および 8 については、それぞれ 7:00-8:00, 3:00-4:00, 12:00-13:00 の区間データが占める割合が多いが、属する区間データの数が少ないため、特定の傾向を見いだすことができなかった。

図 6 より、変動値パラメータの主成分にもとづくクラスタリングでは、多くの区間データはクラスタ 1 か 2 に属していた。しかしながら、クラスタ 1 には曜日によらず 7:00 - 8:00 の区間データがすくなくとも一つは属しており、クラスタ 2 には曜日によらず 3:00 - 4:00 の区間データがすくなくとも一つは属していた。また、クラスタ 1 には日曜日の各計測時間ごとの区間データがすくなくとも一つは属しており、クラスタ 2 には月曜日、水曜日、金曜日、土曜日の各計測時間ごとの区間データがすくなくとも一つは属していた。したがって、区間データの曜日や時間帯に応じてクラスタ 1 か 2 のどちらにより属しやすいかといった傾向はありそうだが、また、クラスタ 3 には水曜日の区間データが一つも属しておらず、クラスタ 4 には金曜日の区間データが一つも属していなかった。

以上より、時系列モデリングによって得られるパラメータにもとづくクラスタリングを用いて異常検知を行う方法としては、実測値パラメータまたは実測値パラメータの主成分によるクラスタリング結果をもとに、時間帯や曜日ごとにモデルパラメータのテンプレートをその傾向を捉えたクラスタの代表点から用意し、その下で回帰予測を行い得られる応答遅延のベースラインの予測値から実測値が大きく外れ続けた場合に異常を検知する手法や、現場の無線端末上でリアルタイムに逐次クラスタリングを行い、その計測データが収容されるクラスタが時間帯や曜日の傾向を捉えたクラスタから外れ続けた場合に異常を検知する手法などが考えられる。

## 5. まとめと今後の課題

本稿では、応答遅延の実測値および変動値のベースラインの時間経過に伴う変動の仕方を ARMA-GARCH モデルのモデルパラメータから捉えることが可能であることを示した。さらに、標準化されたモデルパラメータの第三主成分までを用いてクラスタリングを行うことで、その時間帯や曜日に応じたある傾向を捉えられることが可能であることを示した。これにより、ARMA-GARCH モデルを用いて応答遅延の実測値および変動値のベースラインの変動の仕方を求め、通常の傾向と対比して行う異常検知手法が可能と考える。障害検知手法としては、例

えば、前もって行ったクラスタリング結果において形成された曜日や時間帯ごとの傾向を持つクラスタの代表点を、その曜日や時間帯における ARMA-GARCH モデルのパラメータのテンプレートとし、それを用いて行った応答遅延の予測値から実測値が大きく外れ続けた場合に異常を検知する手法が考えられる。または、現場の無線機器で逐次クラスタリングを行った結果が、その計測時における曜日や時間帯の傾向を持つクラスタに属しなかった場合に異常を検知する手法も考えられる。

今後の課題として、曜日や時間帯ごとの傾向はありそうだがそれは顕著なものではなかったため、クラスタリング結果に基づいた異常検知手法だけでは誤検知や検知漏れが発生すると考えられる。そのため、実用するにあたっては、時系列解析ベースで異常検知を行いながら他の手法と組み合わせる必要があると考える。それには例えば、時系列モデルを用いた回帰分析から大きく外れるものつまりはスパイクの発生頻度や大きさが通常時とは異なる場合に異常を検知する手法などが考えられる。

## 文 献

- [1] A.C. Lima-Filho, R.D. Gomes, M.O. Adissi, T.A.B. Da Silva, F.A. Belo, and M.A. Spohn, "Embedded system integrated into a wireless sensor network for online dynamic torque and efficiency monitoring in induction motors," IEEE ASME transactions on mechatronics, vol.17, no.3, pp.404-414, 2012.
- [2] 辰巳 陸, "実環境における Radio Tomographic Imaging によるイベント検知の検証," 特別研究報告, 大阪大学, Feb 2019.
- [3] A. Striegel, S. Liu, X. Hu, and L. Meng, "Lte and wifi: Experiences with quality and consumption.," FNC/MobiSPC, pp.418-425, 2014.
- [4] J. Garcia, S. Alfredsson, and A. Brunstrom, "Delay metrics and delay characteristics: A study of four swedish hsdpa+ and lte networks," 2015 European Conference on Networks and Communications (EuCNC)IEEE, pp.234-238 2015.
- [5] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z.M. Mao, S. Sen, and O. Spatscheck, "An in-depth study of lte: effect of network protocol and application behavior on performance," ACM SIGCOMM Computer Communication Review, vol.43, no.4, pp.363-374, 2013.
- [6] 田行里衣, 池上 大介, "LTE 環境下の実測データに基づく RTT 分布モデリング," 電子情報通信学会, vol.118, no.140, pp.57-61, 2018.
- [7] C.G. Lamoureux and W.D. Lastrapes, "Persistence in variance, structural change, and the garch model," Journal of Business & Economic Statistics, vol.8, no.2, pp.225-234, 1990.
- [8] H. Bozdogan, "Model selection and akaike's information criterion (aic): The general theory and its analytical extensions," Psychometrika, vol.52, no.3, pp.345-370, 1987.
- [9] K.P. Burnham and D.R. Anderson, "Multimodel inference: understanding aic and bic in model selection," Sociological methods & research, vol.33, no.2, pp.261-304, 2004.
- [10] P.R. Hansen and A. Lunde, "A forecast comparison of volatility models: does anything beat a garch (1, 1)?," Journal of applied econometrics, vol.20, no.7, pp.873-889, 2005.
- [11] I.T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol.374, no.2065, p.20150202, 2016.
- [12] F. Murtagh and P. Legendre, "Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion?," Journal of classification, vol.31, no.3, pp.274-295, 2014.

- [13] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, “Understanding of internal clustering validation measures,” 2010 IEEE International Conference on Data MiningIEEE, pp.911–916 2010.
- [14] Ryosuke Kanajiri, “Anomaly detection using incremental clustering and correlation coefficients in industrial wireless sensor networks,” Master’s thesis, Osaka University, Feb 2020.
- [15] K. Mouratidis, D. Papadias, and S. Papadimitriou, “Medoid queries in large spatial databases,” International Symposium on Spatial and Temporal DatabasesSpringer, pp.55–72 2005.