

SARIMA によるトラフィック予測に関する研究

B-6

Research on Traffic Prediction by SARIMA

小板橋友晃[†] 塩本公平[†]Tomoaki KOITABASHI[†] Kohei SHIOMOTO[†][†] 東京都市大学大学院総合理工学研究科情報専攻[†] Tokyo City University Graduate Division,

Graduate School of Integrative Science and Engineering, Informatics

1. はじめに

IT 技術の発展とともにモバイルトラフィックデータ量は増加し、人の流れやその時のイベントなどの突発的な事象に影響を受け、予測が困難な複雑なものになっている。

そのような状況下においても適切にトラフィックを予測することができれば、予測を基にクラウドサービスのメモリーや CPU といったリソースの割当てを行いサービス品質の向上が期待できる。

本研究では実際のトラフィックのテストデータを用いて SARIMA[1,2]で予測を行う。実際のトラフィックを用いて予測を行うことで実データにおける予測の有効性を検証する。

2. 予測手法

時系列データであるトラフィックデータの振る舞いを説明及び予測するための時系列分析法を用いる。等時間間隔で観測された入力データを参照し、ARIMA や SARIMA などの線形モデルを使用して系列を近似することでデータの将来の値を予測する。

SARIMA は以下の式で表される[1,2]。

$$x_n = \sum_{i=1}^p a_i x_{n-i} + \sum_{i=0}^q b_i \varepsilon_{n-i} + c$$

$$+ \sum_{j=1}^P A_j \sum_{i=1}^p a_i x_{n-sj-i} + \sum_{j=1}^Q B_j \sum_{i=0}^q b_i \varepsilon_{n-sj-i}$$

$$b_0 = 1 \quad (1)$$

ここで P, p は自己回帰(AR: Auto Regressive)モデルのパラメータ、 Q, q は移動平均(MA: Moving Average)モデルのパラメータで、依存するデータとエラーの数を表す。また A, a, B, b は係数、 ε_n は誤差を表す。

入力とするデータは非定常データであり、周期性(季節性)要因が存在するので時系列方向(式(2))と周期方向(式(3))に対し差分演算を行い入力とする。差分演算を行う回数を d, D, s は周期を表す。

$$\Delta^d x = x_n - x_{n-1} \quad (2)$$

$$\Delta^D x_n = x_n - x_{n-s} \quad (3)$$

時系列方向へのパラメータを p, d, q 、季節成分へのパラメータを P, D, Q, s とするとき SARIMA モデルは $SARIMA(p, q, s)(P, D, Q, s)$ と表す。

SARIMA モデルは AIC を用いて評価する。AIC は以下の式で定義される。

$$AIC = -2 \log L + 2k \quad (4)$$

ここで L は最大尤度、 k は $k = p + q + P + Q$ である。この AIC が小さい値をとるほど良いモデルである。

3. 性能評価

SARIMA で実際に予測を行った。入力としたトラフィックデータはネットワーク設計のためのテストインスタンスとし

て公開されている SNDlib[4,5]のライブラリの geant を使って作成した。

本実験では 5 日分の入力データを使い SARIMA モデルを推定し、そのモデルを基に 1 日分のトラフィックの予測を行い予測値と実測値を比較する。

AIC を用いてモデルの評価をした結果、本実験においては $SARIMA(4,1,1)(1,1,1,96)$ に決定した。

図 1 は予測の結果である。横軸が時系列、縦軸がトラフィック量を表している。予測値はトラフィックの増減に関して近い変動を示したが、値そのものは大きく外れていることがわかる。直前の増加傾向を反映してしまっただと考えられる。実際のトラフィックデータにはより急激な変化が観測できると予想できるためこの手法でのトラフィック予測は困難である。

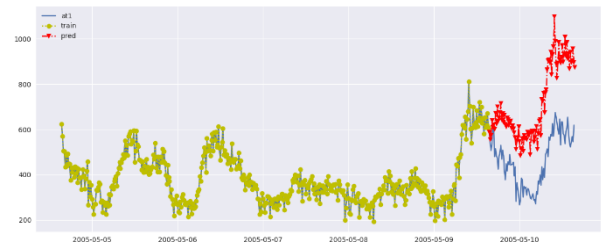


図 1 SARIMA での予測

4. むすび

本論においては SARIMA でのトラフィックの予測の評価を行った。実験結果から SARIMA での予測精度はあまりよくないといえる。トラフィックの複雑化が増す今日においてはさらに困難であると考えられるため、トラフィック以外のデータを用いるなど別のアプローチを検討する必要がある。

参考文献

- [1] S. Medhn, B. Seifu, A. Salem, and D. Hailemariam. Mobile data traffic forecasting in umts networks based on sarima model: The case of addis ababa, ethiopia. In 2017 IEEE AFRICON, pages 285-290, Sep. 2017.
- [2] 馬場 真哉. 時系列分析と状態空間モデルの基礎 R と Stan で学ぶ理論と実装. プレアデス出版, 2018.
- [3] S. Orlowski, M. Pióro, A. Tomaszewski, and R. Wessály. SNDlib 1.0-Survivable Network Design Library. In Proceedings of the 3rd International Network Optimization Conference (INOC 2007), Spa, Belgium, April 2007. <http://sndlib.zib.de>, extended version accepted in Networks, 2009.
- [4] S. Orlowski, M. Pióro, A. Tomaszewski, and R. Wessály. SNDlib 1.0-Survivable Network Design Library. Networks, 55(3):276-286, 2010.