電力使用量の時系列データを用いた予測

小俣満里子　(指導教員：工藤和恵)

1　はじめに

時系列モデルを用いる一つの大きな目的は、過去のデータから将来を予測することである。本研究では自己回帰分析を用いて、使用電力量のデータから使用量を予測することを試みた。

時系列データとは、時間とともに変動する現象に対して時間の順序で測定・観測した結果を記録したデータである。言い換えれば、時系列データは各時点で観測された値の集合である。また、時系列データには周期的に変動する周期性質を持つものも多く、一定の期間で周期的に変動する時系列データについて、データをいくつかの成分に分解することで、より詳細な分析を行うことができることが分かっている。この研究では次のように成分を分解する。

観測値＝トレンド＋周期変動＋残差

ここで、トレンドは比較的滑らかな長期変動を示すものであり、周期変動は、一定の時間あるいは時節とともに一定のリズムで周期的に変動する成分であり、残差はデータから周期変動とトレンドを除去した不規則変動を示す。

2　AR(自己回帰) モデル

時系列モデルを扱う際の根幹にあるのが、定常性という概念である。ここで用いる定常性として、以下に弱定常性の定義を示す [1]。任意の t と k に対して以下が成立する場合、過程は弱定常性という。

$$E(y_t) = \mu \tag{1}$$

$$\mathrm{Cov}(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)] \tag{2}$$

ここで E は期待値、Cov は自己共分散を表す。(2) より、定常過程において、自己共分散は時点には依存せず、時間差 k のみに依存することが分かる。

AR モデルとは、時系列時点 $t-p$ から t までの各データの関係式で、以下の式で表される。

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t \tag{3}$$

ここで a_i $(i = 1, 2, ..., p)$ は自己回帰係数、p は次数、ε_t はホワイトノイズを表す。ここで、ホワイトノイズ ε_t は全ての時点 t において以下の式が成立するものである。

$$E(\varepsilon_t) = 0 \tag{4}$$

$$\gamma_k \equiv E(\varepsilon_t, \varepsilon_{t-k}) = \begin{cases} \sigma^2, k = 0 \\ 0, k \neq 0 \end{cases} \tag{5}$$

ここで、σ^2 は分散を表す。時系列モデルの確率的変動は、ホワイトノイズで表されることが多く、時系列分析で重要になってくる。自己回帰分析では、適切に次数 p を決定することと、自己回帰係数 a_i を推定し、良いモデルを求めることが主な作業である。

3　方法

時系列データを扱うために、R 言語を用いて実際にデータの予測を行なった。データは東京電力の過去の使用電力量データ [2] のうち、2016 年 4 月から 2017 年 3 月のデータを用いた。また、1 時間ごとのデータを 1 日ごとのデータ量に変換して使用した。R 上で 8 月、2 月それぞれの 4 週間分のデータを時系列データとして読み込み、次の 1 週間の予測を行なった。

3.1　モデルの推定

モデルの推定は、自己回帰分析において次数と自己回帰係数を適切に求める作業である。AR モデルの推定する方法には、ユールウォーカー法・最小 2 乗法・最尤法などがあるが、本研究では R 言語でデフォルトで設定されているユールウォーカー法を使用する。また、最適なモデルを選択・評価する客観的な基準として情報量基準があり、R 言語では赤池情報量基準 (AIC) が用いられる。

$$\mathrm{AIC} = -2L(\hat{\theta}) + 2k \tag{6}$$

ここで L は対数尤度を最尤推定した最大尤度、$\hat{\theta}$ はパラメータの予測値、k は推定したパラメータの数を表す。AIC が最小のモデルを選択することが最適なモデルになるとされている [1]。

3.2　モデルの診断

作成したモデルの適切さを判断するには、残差分析が必要となる。AR モデルにおける残差は、平均 0 の正規分布に従い、残差は互いに独立で相関がないことが望ましいとされる [3]。R でのモデルの診断はリュングボックス検定を使用する。データが自己相関関係を有していない（自己相関係数が 0）という帰無仮説を R 言語で検定すると、p 値（帰無仮説が正しい場合の検定統計量の裾確率を評価したもの）が出力される。有意水準 95%とすると、p 値が 0.05 より小さければ帰無仮説を棄却すれば良い。p 値が大きい場合、帰無仮説は棄却できないので自己相関がないと言え、モデルが適切であることが多いとされる。

3.3　モデルの予測

モデルの予測では、実際に推定したモデルを使って値を求める。本研究では実際に 4 週間分のデータから推定したモデルを使い、R 言語の関数 predict を用いて次の 1 週間の値の予測を行なった。

4　結果

本研究では 8 月と 2 月のデータを利用して実験を行なったが、はじめに 2 月の実験結果を示す。以下の式が 2 月のデータをそのまま使用して R 言語で推定した AR モデルである。

$$y_t = 0.7448 y_{t-1} - 0.4173 y_{t-2} \tag{7}$$

診断結果は p 値が 0.8709 となり、悪い推定ではないとされた。上記のモデルを利用した予測結果を実データと比較したものを図 1 に示す。

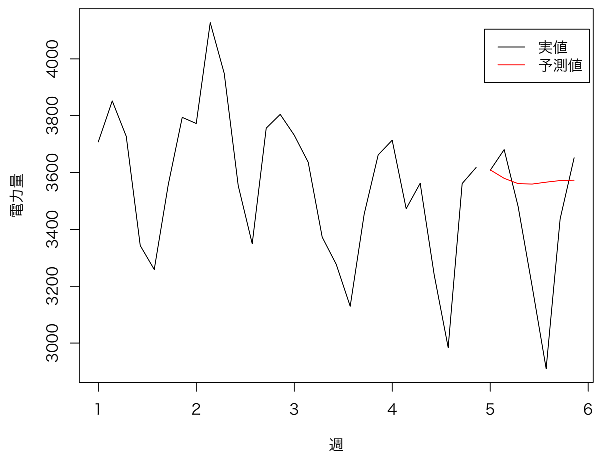

図 1: AR を用いた予測結果 (2 月)

図 1 より、データをそのまま使用し推定した AR モデルでは、診断結果が悪くないにも関わらず、良い予測ができていないことが分かる。そこで、データの成分分解を行なった。結果を図 2 に示す。

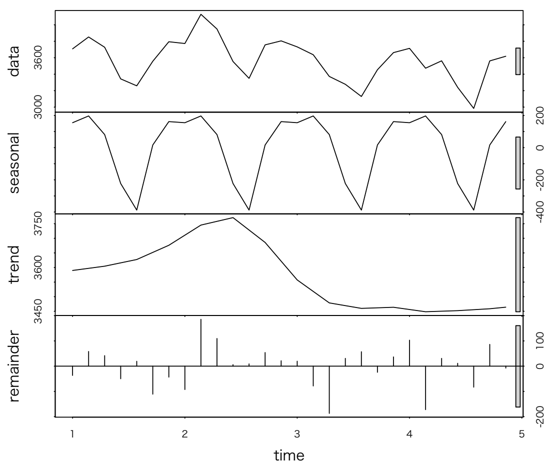

図 2: データの成分分解結果 (2 月)

図 2 で、1 段目は実際のデータの値、2 段目は周期変動、3 段目がトレンド、4 段目が残差を表している。図 2 より、電力データは 1 週間ごとの周期変動があるということがわかる。そこで、元のデータから周期変動を取り除いたものを新たなデータとして AR モデルに当てはめた。その結果が以下の式である。

$$y_t = 0.7486y_{t-1} - 0.5644y_{t-2} + 0.4561y_{t-3} \tag{8}$$

このモデルを診断した結果、p 値が 0.8564 となった。よって、このモデルは自己相関がなく、比較的良い推定であることが言える。このモデルを適用し、予測した結果に周期変動を加えたものを予測値とし、実データと比較したものが図 3 である。同様にして、8 月のデータも周期変動を除いたもので予測を行なった。その結果を図 4 に示す。

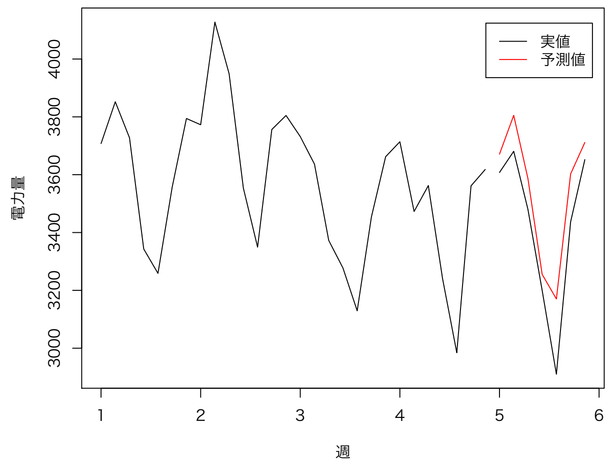

図 3: 周期変動を考慮した AR による予測結果 (2 月)

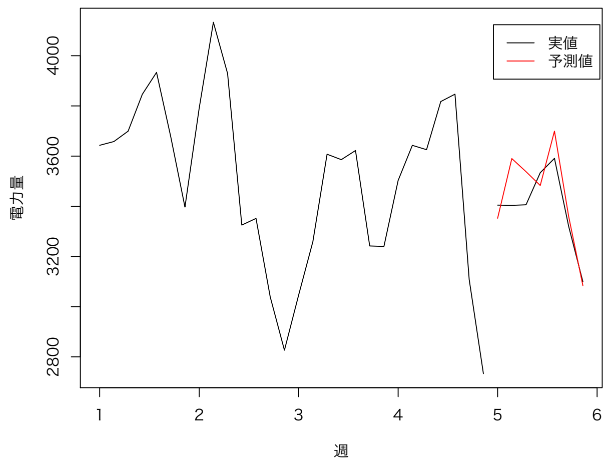

図 4: 周期変動を考慮した AR による予測結果 (8 月)

5 まとめ

時系列データの自己回帰分析による予測を行なった。自己回帰モデルを推定する際に、そのままのデータを使い推定する場合と、周期変動を除いてから行う場合の結果を比べた。その結果、データをそのまま使用した場合にはあまり良い予測ができなかったが、データを分解してモデルを推定した際には、比較的良い予測ができることが分かった。周期変動のある時系列データでは、周期変動を取り除いて自己回帰分析をすることで、AR モデルでも比較的良い予測をすることができる。

参考文献

[1] 沖本竜義, 経済・ファイナンスデータの計量時系列分析, 朝倉書店 (2010)

[2] 東京電力パワーグリッド 過去の電力使用実績データのダウンロード, http://www.tepco.co.jp/forecast/html/download-j.html, (2017 年 12 月参照)

[3] 金明哲, R によるデータサイエンス, 森北出版 (2007)