Vol.2013-MPS-92 No.17
2013/2/28

ベイズ推定を用いた指数忘却型自己回帰モデルによるトレンド，季節性を含むデータの予測 (2013年1月30日版)

飯田 紘士[1,a)]　勝木 孝行[2]　恐神 貴行[2]　中川 裕志[1]

概要：電力需要予測問題のようなトレンドと季節性を含むデータの予測には，単純指数平滑化法を発展させ，データの指数的な忘却を表現できる Holt-Winters 法や discounted least squares を中心に様々な方法が提案されてきた．一方で，モデルが複雑になるに従い，そのパラメータの最適化は，過学習を起こしやすい多変数の同時推定となるため，精度や安定性に問題があることが知られている．特に，従来手法はいくつかのパラメータで過学習の危険の大きい点推定を用いていることが多く，それが精度や安定性の劣化要因となっていた．提案手法では AR model と discounted least squares を組み合わせた表現力の高いモデルを使い，かつ，過学習を適切に防げるよう全てのパラメータをベイズ推定を用いて推定することで，従来手法を上回る予測精度を得ることに成功した．その際効率的な推定と変数の分布の適切な設定のために変分ベイズ法と Taylor 近似を組み合わせた近似を行った．

キーワード：時系列予測，電力需要予測，Discounted Least Squares，指数平滑化法，フルベイズ推定，変分ベイズ，テイラー近似，ベータ分布

Forecasting for Trend and Seasonal Time Series Using Exponential Weighted Baysian Approach (version 2013/1/30)

Abstract: To forecast for trend and seasonal time series, many exponential weighted methods have been proposed such as the Holt-Winters method and regression methods using discounted least squares. However, conventional methods need to estimate many parameters simultaneously by point estimation, which can lead to overfitting. We propose a statistical model based on discounted least squares. In this model, a discount factor and trend/seasonal components are represented as hidden variables, and all the parameters are evaluated as distributions. To estimate these variables, we use Taylor approximations and Variational Bayes. As a result, we solve these problems and our proposed method predicts trend and seasonal time series more accurately than conventional methods.

Keywords: Time Series Forecasting, Load Forecasting, Discounted Least Squares, Exponential Smoothing, Fully Bayes Approach，Variational Bayes，Tayor Approximations, Beta Distribution

1. はじめに

過去の観測値のみを用いて，トレンドと季節性を含むデータの予測を行う方法は電力需要予測の分野を中心に数多く研究されている．これらのデータを解析する際には，トレンドと季節性を表現できるモデルを用いることが好ましく，また，時系列データの性質上，過去よりも現在のデータを重視した予測を行えることが望ましい．過去のデータよりも現在のデータを重視する方法論としては，指数平滑化法 (exponential smoothing) や discounted least squares (DLS) が代表的である．この 2 つの手法はもっとも単純な

[1] 東京大学大学院 情報理工学系研究科
Graduate School of Informatinal Science and Techonology, The University of Tokyo
[2] 日本アイ・ビー・エム（株）東京基礎研究所
IBM Research - Tokyo
[a)] Hiroshi_Iida@mist.i.u-tokyo.ac.jp

Vol.2013-MPS-92 No.17
2013/2/28

形の時，どちらも単純指数平滑化法と呼ばれる同じ式で表される．しかしトレンドや季節性を考慮する際はアプローチの仕方が大きく異なる．

指数平滑化法は，予測式に着目し，予測式内のトレンドや季節性を表す項のそれぞれを個別に重み付けすることで，新しく観測されたデータをより重視した予測を行えるようなモデルを構築する．この方法は Holt-Winters 法と呼ばれ，時系列予測の分野でよく用いられている．Holt-Winters 法は複雑なモデルを扱うため，パラメータの推定が難しい．そのため推定の際には，複数の平滑化係数を grid search して，予測値と観測値の平均二乗誤差を最小にするパラメータを採用することが多い．

一方 DLS では，誤差関数に着目し，過去のデータ程現在の予測関数との当てはまりが悪いとする自然な誤差関数を考えて予測関数の最適化を行う．具体的には，まず予測に用いる関数 $f(t)$ と平滑化係数 γ を用意する．そして，時刻 T までのデータを観測した時，各時刻 t における予測値 $f(t)$ と観測値 X_t の誤差に指数的な重みを付けた二乗誤差

$$\sum_{i=1}^{T} \gamma^{T-i}(X_i - f(i))^2 \tag{1}$$

の最小化により $f(t)$ に含まれるパラメータの推定を行う．この誤差関数を用いた場合，パラメータ推定の際に，自然に過去のデータによる影響を割り引くことが可能になる．しかし，DLS を用いた従来研究ではトレンドや季節性を同時に表現するような表現力の高いモデルを使った手法は提案されてこなかった．

本論文では，DLS に基づく統計的なモデルを用いて予測を行う．特に，DLS の忘却の仕組みに着目し，従来モデルの欠点である表現力の低さを，トレンドと季節性を同時に表現可能な $f(i)$ を用いることで解決する．さらに，提案モデルを用いて $f(i)$ のパラメータを推定するには，過学習の危険の大きい多変数の同時最適化が必要になるが，全てのパラメータに確率分布を仮定したフルベイズ推定を行うことで，過学習を適切に防ぎ，安定かつ高精度な推定を実現する．平滑化係数 γ は忘却を適切に表現するため、$0 \le \gamma \le 1$ であることが望ましい．この定義域を満たすため，γ の事後分布がベータ分布となるよう，Taylor 近似と変分ベイズ法を組み合わせた近似推定を行うのが提案手法の一つの特徴である．本論文の貢献は以下の 2 つである．

- 表現力の高いモデルを用いて，過学習を適切に防ぐことが出来るフルベイズ推定を行うことで，実データを用いた実験において，Holt-Winters 法よりも優れた予測精度を達成した．
- 平滑化係数を推定するために，変分ベイズ法に基づいたアルゴリズムを提案した．これにより平滑化係数を決定するための確立した手法がない中で，1 つの解決策を与えた．

表 1 記号の定義
Table 1 notations

Term	Meaning
$\mathcal{N}(x;\mu,\sigma^2)$	$(2\pi\sigma^2)^{-\frac{1}{2}}e^{(-\frac{1}{2\sigma^2}(x-\mu)^2)}$
$\mathrm{Beta}(x;a,b)$	$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}x^{a-1}(1-x)^{b-1}(0 \le x \le 1)$
$\mathrm{Gamma}(x;a,b)$	$\frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx}(x>0)$
X_i	時刻 i における観測値
$X_{d,h}$	観測を始めて d 日と h 時間後の観測値 $(1 \le h \le 24)$
X	観測変数の集合
D	訓練データの日数の合計
γ	指数的な忘却を表す平滑化係数 $(0 \le \gamma \le 1)$
μ	データの線形なトレンドを表す確率変数 $(\mu \in \mathbb{R})$
θ	一時間前の観測値に対する自己回帰係数 $(\theta \in \mathbb{R})$
Y_h	データの一日周期の変動を表す確率変数 $(Y_h \in \mathbb{R})$
α	Y_h が従う正規分布の精度 $(\alpha > 0)$
β	X_i または $X_{d,h}$ が従う正規分布の精度 $(\beta > 0)$
$f(t)$	DLS で時刻 t における予測値を出力する関数

表 1 には，記号および変数の定義をまとめた．

2. 関連研究

トレンドとはデータの長期的な増加または減少傾向の事を指す．また今回扱う季節性は，一日周期の各時刻によるデータの変動とする．DLS で予測に用いる $f(t)$ の中に季節性を表す項として三角関数を用いる [3] が提案されている．DLS において (1) 式の γ は通常 discount factor と呼ばれる．しかし，指数的な忘却率を表すという点で，指数平滑化法の平滑化係数と同様であり，本稿ではどちらも平滑化係数と呼ぶ．

DLS は予測関数に単純な関数を用いることが多く，表現力が低い事が指摘されてきた．トレンドと季節性を含むデータに対して指数的な忘却を行う方法は [7] が詳しい．[7] では指数平滑化法や DLS 及びその他の手法を扱っている．提案手法では定数項付きの AR model と季節性を表す項に DLS を組み合わせる．AR model に DLS を組み合わせる方法は [4] で行われているが，[4] では季節性を考慮していないことに加え，平滑化係数を推定していない．提案手法の季節性を表す項の表現は [6] の方法を用いているが，[6] ではデータの指数的な忘却を行っていない．今回提案するモデルは DLS に基づいているが，データの指数的な忘却を行う方法の中で非常に良く研究されてきた指数平滑化法と比較を行う．指数平滑化法への統計的なアプローチは [5] を基に研究がなされてきた．特にパラメータに分布を仮定するベイズ推定は [1]，[2] などで行われている．これらは Holt-Winters 法を対象としている点で異なる．

3. 提案するモデル

3.1 問題の設定

時刻 1 から時刻 $t-1$ までの時系列データ

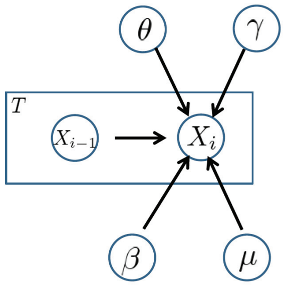

図 1 トレンドから成る時系列を扱うモデル
Fig. 1 Graphical Model with Trend

$X_1, X_2, \cdots, X_{t-1}$ が与えられた時に，時刻 t での実現値 X_t を予測したい．本論文で実験に用いる時系列データ X は，後述する線形なトレンドを表す人工データと，実際の電力需要のデータである．予測は，$X_1, X_2, \cdots, X_{t-1}$ が与えられた上で，$f(t)$ 内の全パラメータの事後分布を導出し，その事後分布によって $f(t)$ の期待値をとることで，予測値とする．以降の節で DLS の誤差関数を，確率モデルを用いて再定義し，フルベイズ推定の枠組みで $f(t)$ のパラメータの事後分布を求める．

3.2 DLS の確率モデルによる表現

過去のデータが従う正規分布の精度を減衰させるモデルを用いることで，DLS のモデルを表現出来る．この事を，単純指数平滑化法の例を通して説明する．なお，後述の変分ベイズ法による推定がうまく行えるように，提案するモデルではすべて $1-\gamma$ のべき乗によってデータの忘却を行う．単純指数平滑化法では時刻 T までの指数移動平均

$$a_T = \gamma \sum_{i=1}^{T} (1-\gamma)^{T-i} X_i \tag{2}$$

を用いて予測を行う．時刻 $T+1$ での予測値は a_T である．a_T は過去のデータを指数的に減衰させながら重みづけした値となっている．DLS で単純指数平滑化法を行うためには，平滑化係数を $1-\gamma$，予測に用いる関数を $f(t) = a$ とすれば良い．時刻 T までのデータを観測した時，DLS の目的関数は，(1) 式より，

$$\min_a \sum_{i=1}^{T} (1-\gamma)^{T-i} (X_i - a)^2 \tag{3}$$

である．この目的関数は，時刻 i のデータが平均 a，精度 $(1-\gamma)^{T-i}\beta$ の正規分布から生成されるというモデルの最尤推定として得ることもでき，この場合の，$X \equiv (X_1, \ldots, X_T)$ の同時分布の確率密度関数は，

$$p(X|a) = \prod_{i=1}^{T} \mathcal{N}(X_i; a, ((1-\gamma)^{T-i}\beta)^{-1}) \tag{4}$$

である．(4) 式の対数尤度を a について最大化する最適化問題は，

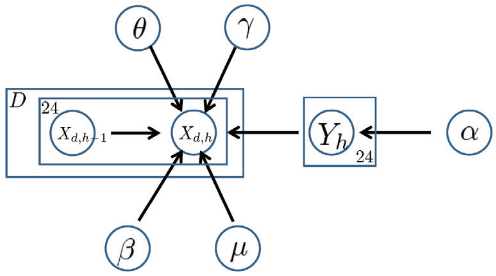

図 2 トレンドと季節性を含む時系列を扱うモデル
Fig. 2 Graphical Model with Trend and Seasonality

$$\max_a -\frac{\beta}{2} \sum_{i=1}^{T} (1-\gamma)^{T-i} (X_i - a)^2 \tag{5}$$

となり，(3) 式と一致する．これを解くと，

$$a = \frac{\sum_{i=1}^{T} (1-\gamma)^{T-i} X_i}{\sum_{i=1}^{T} (1-\gamma)^{T-i}} \tag{6}$$

を得る．学習するデータの数 T が十分大きいとして $(1-\gamma)^T = 0$ とみなすと，

$$a = \gamma \sum_{i=1}^{T} (1-\gamma)^{T-i} X_i \tag{7}$$

であり，これは単純指数平滑化法である．同様の議論により，次の性質が成り立つ．

- 平均が $f(t)$，精度が $(1-\gamma)^{T-i}\beta$ の正規分布により X_i が生成されると仮定する．$f(t)$ に含まれるパラメータの最尤推定は，平滑化係数を $1-\gamma$，予測に用いる関数を $f(t)$ として DLS を行った時の目的関数の最適化と等価である．

3.3 トレンドから成る時系列を扱うモデル

この項では線形なトレンドから成る時系列を扱うためのモデルを提案する．このモデルはグラフィカルモデルを用いて図 1 のように表すことができる．時刻 T までのデータを観測した時，時刻 i のデータは

$$p(X_i \mid X_{i-1}, \gamma, \mu, \theta, \beta) \equiv$$
$$\mathcal{N}(X_i \mid \mu + \theta X_{i-1}\ ,\ ((1-\gamma)^{(T-i)}\beta)^{-1}), \tag{8}$$
$$p(\mu; \mu_0, \tau_0) \equiv \mathcal{N}(\mu; \mu_0, \tau_0), \tag{9}$$
$$p(\theta; \mu_1, \tau_1) \equiv \mathcal{N}(\theta; \mu_1, \tau_1), \tag{10}$$
$$p(\beta; a_1, b_1) \equiv \mathrm{Gamma}(\beta; a_1, b_1), \tag{11}$$
$$p(\gamma; a_2, b_2) \equiv \mathrm{Beta}(\gamma; a_2, b_2), \tag{12}$$

に従って生成されると仮定する．μ_0，μ_1，τ_0，τ_1，a_1，a_2，b_1，b_2 はハイパーパラメータである．推定の中心となる式は，X_i が平均が $\mu + \theta X_{i-1}$ で精度が $(1-\gamma)^{(T-i)}\beta$ の正規分布から生成されるという式である．先ほどと同様に μ, θ に関して最尤推定を行うと，$f(t) = \mu + \theta X_{t-1}$ に対して DLS を用いた時の目的関数と一致する．上で仮定した事

前分布は γ の分布以外は全て共役事前分布となっている．ここで，定数項付きの AR model はデータの線形なトレンドを表現することができる．具体的には，線形のトレンド $a+bt$ から成るデータに対して，定数項 μ が傾き b を表すことが知られている．これらのパラメータの推定方法は 4 節で，線形のトレンドから成るデータに対する実験結果は 5 節でそれぞれ後述する．

3.4 トレンドと季節性を含む時系列を扱うモデル

この項では電力需要予測に用いるモデルを提案する．このモデルはグラフィカルモデルを用いて図 2 のようにあらわすことができる．データはトレンドと一日周期の季節性を含み，一時間刻みで与えられていると仮定する．式で表すと，

$$p(X_{d,h} \mid X_{d,h}, \gamma, \mu, Y_h, \theta, \beta) \equiv \mathcal{N}(X_{d,h} \mid \mu + Y_h + \theta X_{d,h-1},\ ((1-\gamma)^{(D-d)}\beta)^{-1}), \quad (13)$$

$$p(Y_h \mid \alpha) \equiv \mathcal{N}(Y_h \mid 0\ ,\ \alpha^{-1}), \quad (14)$$

$$p(\mu; \mu_0, \tau_0) \equiv \mathcal{N}(\mu; \mu_0, \tau_0), \quad (15)$$

$$p(\theta; \mu_1; \tau_1) \equiv \mathcal{N}(\theta; \mu_1, \tau_1), \quad (16)$$

$$p(\alpha; a_0, b_0) \equiv \mathrm{Gamma}(\alpha; a_0, b_0), \quad (17)$$

$$p(\beta; a_1, b_1) \equiv \mathrm{Gamma}(\beta; a_1, b_1), \quad (18)$$

$$p(\gamma; a_2, b_2) \equiv \mathrm{Beta}(\gamma; a_2, b_2), \quad (19)$$

というモデルである．このモデルでは，前項のモデルに季節性を表す項として，Y_h を加えた．Y_h が従う正規分布の精度にはさらに，ガンマ分布に従う確率変数 α を仮定した．このモデルでは新しいデータがくるたびにパラメータを学習すると，計算時間がとても長くかかってしまうので，パラメータの更新を一日に一度行えば良い様に，一日ごとにデータへの重みづけを減衰させるという設定にしている．本稿は一つの季節性のみを含むデータを対象としているが，一週間ごとの曜日による周期などを同様にして表す事ができる．

4. 事後分布の推定と予測について

4.1 変分ベイズ法

3.3 項，3.4 項のモデルの事後分布を解析的に計算するのは困難である．この事後分布の推定を行うために変分ベイズ法を用いた．以下では観測変数の集合を X，確率変数の集合を Z とする．変分ベイズ法では，設定した分布 $q(Z)$ の中から最もよく $p(Z|X)$ を近似するものを探索する．変分ベイズ法では一般的に，各々の確率変数が独立な確率分布に従うとして，それらの積を $q(Z)$ とする．即ち，確率変数が M 個あるとすると $q(Z)=\prod_{i=1}^{M} q_i(Z_i)$ が成立していると仮定した．変分ベイズ法では最適な $q^*(Z)$ を，$p(Z|X)$ との KL 情報量を最小化する

$$q^*(Z) = \underset{q(z)}{\operatorname{argmin}}\ D_{KL}\ (q(Z)\ ||\ p(Z|X)) \quad (20)$$

とする．(20) 式は，解析的に解くことが出来ない．そこで，

$$\ln p(X) = \int q(Z) \ln \frac{p(X,Z)}{q(Z)} \mathrm{d}Z + D_{KL}(q(Z)\ ||\ p(Z|X)) \quad (21)$$

の成立を利用する．(21) 式により，KL 情報量の最小化は

$$\int q(Z) \ln \frac{p(X,Z)}{q(Z)} \mathrm{d}Z \quad (22)$$

の最大化と等価であり，実際の最適化を行う際には (22) 式を用いる．この最適化は各分布 q_j を順番に更新していく方法で行う事が出来る．各更新において q_j^* は，

$$q_j^*(Z_j) \propto \exp(\mathbb{E}_{\prod_{i\neq j} q_i(Z_i)}[\ln p(X,Z)]) \quad (23)$$

であることを利用して更新を行う．

4.2 事後分布の推定について

この項では，3.4 項で提案したモデルの事後分布を推定する方法を説明する．3.3 項のモデルについても同様の推定を行う事が出来るため，ここでは省略する．今回用いた設定では，$Z = (\gamma, \mu, \theta, Y_h, \alpha, \beta)$ および，$q(Z) = q_1(\gamma)q_2(\mu)q_3(\theta)q_4(Y_h)q_5(\alpha)q_6(\beta)$ とした．

提案したモデルにおいて，γ 以外の変数に関しては，事前分布として共役な事前分布を用いているため，(23) 式により各分布の更新を行えば良い．しかし $q_1(\gamma)$ の扱いについては少し注意をしなければならない．$q_1^*(\gamma)$ は

$$q_1^*(\gamma) \propto \exp(\mathbb{E}_{q(\mu,\theta,Y_h,\alpha,\beta)}[\ln p(X,Z)]), \quad (24)$$

$$\propto \exp(\mathbb{E}_{q(\mu,\theta,Y_h,\alpha,\beta)}[\ln p(X,\gamma,\mu,Y_h,\theta,\beta) + \ln p(\gamma)]), \quad (25)$$

であるが，上式は共役な形ではないため，$q_1(\gamma)$ の期待値を計算することは困難である．よって，$\exp(\mathbb{E}_{q(\mu,\theta,Y_h,\alpha,\beta)}[\ln p(X,\gamma,\mu,Y_h,\theta,\beta)])$ がベータ分布に共役となるよう近似する．具体的には，$\tilde{\gamma} = \mathbb{E}_{q(\gamma)}[\gamma]$ なる $\tilde{\gamma}$ を用いて，(25) 式に含まれる $(1-\gamma)^n$ を，$\ln\gamma = \ln\tilde{\gamma}$ を中心に一次の Taylor 展開をすることで $\ln\gamma$ の一次式に近似する．$f(\ln\gamma) = (1-\gamma)^n$ とおくと，Talor 展開の式は,

$$f(\ln\gamma) \simeq f(\ln\gamma)|_{\ln\gamma=\ln\tilde{\gamma}} + f^{(1)}(\ln\gamma)|_{\ln\gamma=\ln\tilde{\gamma}}(\ln\gamma - \ln\tilde{\gamma})$$

となるため,

$$(1-\gamma)^n \simeq (1-\tilde{\gamma})^n + \left.\frac{\partial(1-\gamma)^n}{\partial \ln\gamma}\right|_{\ln\gamma=\ln\tilde{\gamma}} (\ln\gamma - \ln\tilde{\gamma}) \quad (26)$$

$$\simeq (1-\tilde{\gamma})^n + n(1-\tilde{\gamma})^{n-1}(-\tilde{\gamma})(\ln\gamma - \ln\tilde{\gamma})$$

となる．この近似式を用いると，

$$q_1^*(\gamma) \propto \exp(\mathbb{E}_{q(\mu,\theta,Y_h,\alpha,\beta)}[\frac{24}{2}\sum_{d=1}^{D}(D-d)\ln(1-\gamma)$$
$$+\frac{\beta}{2}\ln\gamma\sum_{d=1}^{D}\tilde{\gamma}(D-d)(1-\tilde{\gamma})^{D-d-1}$$
$$\times\sum_{h=1}^{24}(X_{d,h}-\mu-Y_h-\theta X_{d,h-1})^2$$
$$+(a_2-1)\ln\gamma+(b_2-1)\ln(1-\gamma)]) \quad (27)$$
$$\propto \gamma^{\hat{a}-1}(1-\gamma)^{\hat{b}-1} \quad (28)$$
$$\propto \mathrm{Beta}(\gamma;\hat{a}_{(\gamma)},\hat{b}_{(\gamma)}) \quad (29)$$

となるため，$q_1^*(\gamma)$ は Beta 分布になる．これにより，$q_1(\gamma)$ の期待値を計算することが可能になる．ここで，$\hat{a},\hat{b}$ はそれぞれ

$$\hat{a}_{(\gamma)} = a_2+\mathbb{E}_{q(\mu,\theta,Y_h,\alpha,\beta)}[\frac{\beta}{2}\tilde{\gamma}\sum_{d=1}^{D}(D-d)(1-\tilde{\gamma})^{D-d-1}$$
$$\times\sum_{h=1}^{24}(X_{d,h}-\mu-Y_h-\theta X_{d,h-1})^2], \quad (30)$$
$$\hat{b}_{(\gamma)} = b_2+\frac{24}{2}\sum_{d=1}^{D}(D-d) \quad (31)$$

である．γ 以外の更新式はここでは割愛する．

5. 実験結果

5.1 目的

3.3 項と 3.4 項で提案したモデルを基に 4 節で説明した方法を用いて予測を行う．適切な忘却率を推定しながら，適切な予測が行えるかどうかを実験により確かめる．

5.2 トレンドから成る人工の時系列での実験

この項では 3.3 項で述べたモデルで，線形のトレンドから成る時系列

$$x_t = a + bt + \epsilon_t \quad (32)$$

の推定を人工データにより実験する．この実験は，3.4 項のモデルを単純にした時に正しくトレンドを推定できるかを確かめるものである．人工データは $a = 5.0$，$b = 1.5$ とした．ϵ_t は，平均 0.0，分散 1.0 の正規分布に従うとして $0 \leq t \leq 999$ の範囲で 1000 点のデータを作成した．まず $0 \leq t \leq 599$ のデータを用いて学習を行う．残りの 400 点では予測を行い，t が一つ増えるごとにパラメータを更新して学習も行った．推定した 400 点でのパラメータの平均は $\mu = 1.48674$，$\theta = 1.00002$，$\gamma = 7.76 \times 10^{-5}$ となり，μ の値がトレンドの傾きを正しく推定している結果が得られた．また，γ はほとんど 0 であり，この場合では忘却はほとんど行われていない．

5.3 電力需要の実際のデータを用いた実験（設定）

本節では，東京電力が公開している，2011 年の一時間

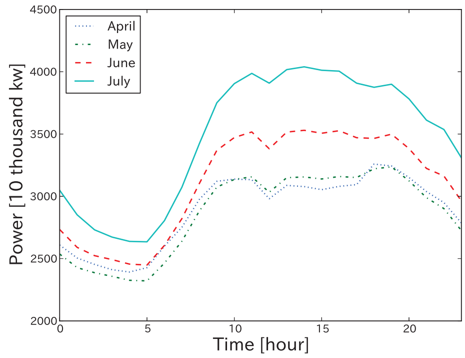

図 3 時刻と電力需要の関係 (夏)
Fig. 3 Comparison between 4 months Load Curves(Summer)

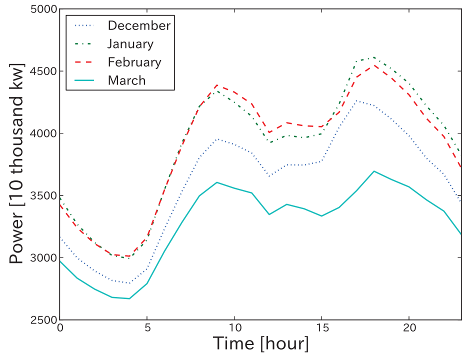

図 4 時刻と電力需要の関係 (冬)
Fig. 4 Comparison between 4 months Load Curves(Winter)

ごとの電力需要データ [*1] を用いて実験を行う．実験では，2011/7/10 0:00〜 2011/7/31 23:00 までを予測する．過去のデータの忘却がどの程度行われるかを理解するために，学習に用いるデータを変えながら実験を行った．具体的には 2011/7/3 0:00 以降の一週間分のデータを学習に用いる設定から，2010/12/15 0:00 以降のおよそ 7 か月分のデータを学習に用いる設定までを実験した．図 3，4 は各月ごとの電力需要の平均を各時刻についてプロットしたものである．この図を見ると，電力需要は 12 月から 3 月までの冬の時期と，4 月から 7 月までの夏の時期で大きく性質が異なることがわかる．冬のデータでは午前 9 時に需要のピークを迎え，日中は朝と夜のピークに比べて需要が落ち込む．夏のデータでは午前 11 時に需要のピークを迎え，日中も冬に比べて需要がそれほど落ち込まない．また夏のデータから読み取れることとして，予測を行う 7 月のデータが 6 月に比べ電力需要がかなり大きいことがわかる．実験では，3.4 項で提案した方法，3.4 項の方法で指数的な忘却を行わない方法，Holt-Winters 法の 3 通りについて比較実験を行う．

最後にそれぞれの手法の設定を説明する．Holt-Winters 法は，一時間分のデータを予測して，新しいデータを得る

*1 http://www.tepco.co.jp/forecast/html/download-j.html

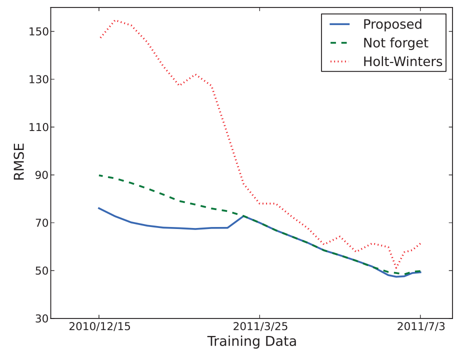

図 5　学習を開始する時期と RMSE の関係
Fig. 5　Training Data vs Result of RMSE

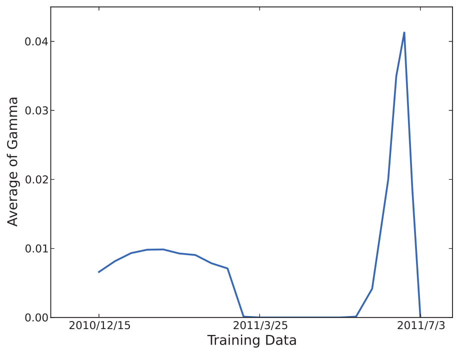

図 6　学習を開始する時期と平滑化係数の関係
Fig. 6　Training Data vs Parameter Estimation Result of γ

度に学習をやり直してパラメータを更新した．提案するモデルでは，一日の始めである 0:00 に毎日学習をやり直してパラメータを更新した．それ以外の時間では，θX_{i-1} の項でデータ自体は用いるが，パラメータの更新は行わない．図 5 では，学習を開始した時期を横軸に，その時の平均二乗誤差 (RMSE: root mean square error) を縦軸にプロットした．また図 6 では，学習を開始した時期を横軸に，縦軸に平滑化係数の平均の値をプロットした．

5.4　電力需要の実際のデータを用いた実験（考察）

実験結果の考察にうつる．提案モデルは 3 つの手法において，どの期間学習したとしても最も良い予測を記録している．まずは提案手法の考察を行う．図 6 を見ると，6 月近辺から学習を開始した時と 3 月以前から学習を開始した時に忘却を積極的に行う．これは前の項で考察したデータセットの性質を良く捉えている．強調したいのは 3 月以前のデータを学習した際の結果である．忘却による効果は図 5 に現れていて，忘却を行わない場合に比べて RMSE が減少していることがわかる．提案手法では $1-\gamma$ のべき乗に従って過去のデータを忘却していくので，冬の時期から学習を開始した場合はおよそ 0.99 のべき乗によりデータを忘却している．一見これではデータをほとんど忘却しないように見えるが，0.99^{35} がおよそ 0.7 で，0.99^{70} がおよそ 0.5，0.99^{120} がおよそ 0.3 であることと，予測を 7 月に行うことから，4 ヶ月前である 3 月近辺のデータは，7 月のデータに比べて 3 割程度の比重しかかけられていない．また 3 月以前のデータはそれよりもさらに小さい重みづけとなっていることがわかる．次に Holt-Winters 法と比較する．Holt-Winters 法との違いも，2011/3/25 以前のデータを含めて学習した時に顕著に見られる．Holt-winters 法では，過去のデータを含めて学習を行うと悪い影響を大きく受けてしまう．それは，Holt-Winters 法では外れ値に影響を受けやすい二乗誤差の最小化によってパラメータを推定するためである．一方で提案手法では，時刻 i のデータは推定した γ に対して $\gamma^{T-i}(X_i - f(i))^2$ という減衰した値でしか影響しないため，精度の低下を招きにくい．

6.　おわりに

提案したモデルを用いて，実データに対して適切な忘却に基づく予測が行えることが確認できた．Holt-Winters 法や DLS は非常によく研究されているが，ベイズ推定を行う試みはあまり多くない．その原因はパラメータの推定が難しいことにある．提案手法では DLS に対して Taylor 近似と変分ベイズ法を用いる事で平滑化係数の推定を可能にした．その際，DLS を時間の経過と共に精度が減衰する正規分布を用いて統計的に表現した．また，DLS を用いるとモデルの表現力が低くなりやすい問題を，トレンドと季節性を表す確率変数を組み合わせる事で解決した．

参考文献

[1] Andrawis, R.R. and Atiya, A.F.: A New Bayesian Formulation for Holts Exponential Smoothing. *Journal of Forecasting*, Vol.28, pp.218-234 (2009)
[2] Bermudez, J.D., Segura, J.V. and Vercheri, E.: Bayesian Forecasting with The Holt Winters Model, *Journal of the Operational Research Society*, Vol.61, pp.164 -171 (2010)
[3] Christiaanse, W.: Short-Term Load Forecasting Using General Exponential Smoothing, *IEEE Transactions on Power Apparatus and Systems* , Vol.90, No.2, pp.900 - 910 (1971).
[4] Jantana, P. and Sudasna-na-Ayudthya, P.: Least Squares and Discounted Least Squares in Autoregressive Process, *Silpakorn University International Journal*, Vol.6, pp.122-135 (2006).
[5] Ord, J. K., Koehler, A. B. and Snyder, R. D.: Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models. *Journal of the American Statistical Association*, Vol.92, pp.1621-1629 (1997).
[6] Reddy, P. and Veloso, M.: Factored Models for Multiscale Decision Making in Smart Grid Customers. *Proceedings of the 26th AAAI* (2012).
[7] Taylor, J. W.: Exponentially Weighted Methods for Forecasting Intraday Time Series with Multiple Seasonal Cycles. *International Journal of Forecasting*, Vol.26, pp.627-646 (2010).