State Space Model for the Prediction of Energy Consumption

Dexiong Chen, Chia-Man Hung

March 1, 2017
Some statistics on the energy consumptions of buildings:

- 48% of the total energy consumption
- 36% of the total CO\textsubscript{2} emission
- 27% of the electric consumption

**Figure:** Energy consumption
Various approaches:
- physics methods
- statistical methods
- neural networks (ANNs)
- support vector machines (SVMs)
- gray model

State space model: simple but extensively used, with theoretical guarantees and confidence intervals.
1. State Space Model and Kalman Filter

2. EM Algorithm

3. Data Challenge: Prediction of energy consumption
the error terms $\varepsilon_t \sim \mathcal{N}(0, P_t)$ and $\eta_t \sim \mathcal{N}(0, Q_t)$ are two independent vector-valued i.i.d. Gaussian sequences.

- $X$ represents state vector sequence, which is not observable.
- $X_0$ is assumed to be $\mathcal{N}(\mu_0, \Sigma_0)$, which is independent of $\varepsilon_t$ and $\eta_t$.
- $\theta_t = (A_t, B_t, C_t, D_t, P_t, Q_t, \mu_0, \Sigma_0)$ are parameters of the model.
Kalman Filter
scheme

Figure: Scheme for Kalman filter
\[ \hat{X}_{t|s} := \mathbb{E}[X_t | Y_{0:s}] \]
\[ \Sigma_{t|s} := \mathbb{E}[(X_t - \hat{X}_{t|s})(X_t - \hat{X}_{t|s})^T] = \text{Cov}(X_t - \hat{X}_{t|s}) \]

By induction, we obtain

\[ \hat{X}_{t|t-1} = A_t \hat{X}_{t-1|t-1} + B_t U_t \] (2)
\[ \Sigma_{t|t-1} = A_t \Sigma_{t-1|t-1} A_t^T + P_t \] (3)
\[ K_t = \Sigma_{t|t-1} C_t^T [C_t \Sigma_{t|t-1} C_t^T + Q_t]^{-1} \] (4)
\[ \hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t (Y_t - C_t \hat{X}_{t|t-1} - D_t U_t) \] (5)
\[ \Sigma_{t|t} = [I - K_t C_t] \Sigma_{t|t-1} \] (6)
We intend to compute $\hat{X}_{t|n}$ et $\Sigma_{t|n}$ via a backward recursion.

$$J_t = \Sigma_{t|t}A_t^T\Sigma_{t+1|t}^{-1}$$  \(7\)

$$\hat{X}_{t|n} = \hat{X}_{t|t} + J_t(\hat{X}_{t+1|n} - \hat{X}_{t+1|t})$$  \(8\)

$$\Sigma_{t|n} = \Sigma_{t|t} + J_t(\Sigma_{t+1|n} - \Sigma_{t+1|t})J_t^T$$  \(9\)

By analogy, we can compute the one-lag covariance

$$\Sigma_{t,t-1|n} = \Sigma_{t|n}J_{t-1}^T$$  \(10\)

These estimates about hidden states will be used in the E-step of the EM algorithm.
1. State Space Model and Kalman Filter

2. EM Algorithm

3. Data Challenge: Prediction of energy consumption
EM algorithm allows to estimate the parameters while some hidden variables present in the model.

\[ L(\theta) = \int f(x; \theta) \lambda(dx) \quad \ell(\theta) = \log L(\theta) \quad (11) \]

The main idea is to consider an intermediate quantity, referred to as a conditional expectation.

\[ Q(\theta; \theta') = \int \log f(x; \theta)p(x; \theta') \lambda(dx) \quad (12) \]
EM Algorithm

We rewrite \( Q \) as

\[
Q(\theta; \theta') = \ell(\theta) - H(\theta; \theta'),
\]

(13)

where

\[
H(\theta; \theta') = - \int \log p(x; \theta)p(x; \theta') \lambda(dx)
\]

(14)

By observing that \( H \) is the Kullback-Leibler divergence, we have the following theorem

**Theoreme**

*Under some assumptions, for any \((\theta, \theta') \in \Theta^2\), we have*

\[
\ell(\theta) - \ell(\theta') \geq Q(\theta; \theta') - Q(\theta'; \theta')
\]

(15)
EM algorithm:

- Expectation step (E-step): compute $Q(\theta; \theta')$.
- Maximization step (M-step): choose $\theta^{i+1}$ to maximize the immediate $\theta^{i+1} = \arg\max_{\theta \in \Theta} Q(\theta; \theta^i)$.

We define here

$$Q(\theta; \theta^i) = \mathbb{E}[\log P(X_{0:n}, Y_{0:n}; \theta) | Y_{0:n}, \theta^i]$$

(16)

Then we can compute explicitly $Q(\theta; \theta^i)$ and its derivatives with respect to each parameter, then express $\theta^{i+1}$ as a function of $\theta^i$. 

Dexiong Chen, Chia-Man Hung
1. State Space Model and Kalman Filter

2. EM Algorithm

3. Data Challenge: Prediction of energy consumption
\[
\begin{align*}
X_{t+1} &= A_t X_t + B_t U_t + \varepsilon_t \\
Y_t &= C_t X_t + D_t U_t + \eta_t
\end{align*}
\]  

\[ (17) \]

- \( U_t \): exogenous data (temperature, sunshine etc)
- \( X_t \): hidden chain
- \( Y_t \): energy use measured by sensors

Missing values in \( U \): prefilled by using a state space model.
Strategy 1: One Regime Model

We apply directly the model to the problem

\[
\begin{align*}
    X_{t+1} &= AX_t + \varepsilon \\
    Y_t &= CX_t + DU_t + \eta
\end{align*}
\]  

(18)
Strategy 1: One Regime Model

Figure: Building 4
We divided states into different groups according to their regime. It is a particular case of time-variant state-space model.

- An instance choice of regimes: “night”, “day”, “day-to-night”, “night-to-day” and “weekend”.
- Remark: when applying the EM algorithm for this model, we compute the empirical average respectively over each regime period to estimate parameters belonging to this regime.
Figure: Building 4
Strategy 3: Two-lag Multi-Regime Model

We replace $U_t$ by $\tilde{U}_t = (U_t, U_{t-1})$ in the previous model,

$$
\begin{align*}
X_{t+1} &= A_{\sigma(t)}X_t + \varepsilon \\
Y_t &= C_{\eta(t)}X_t + D_{\eta(t)}\tilde{U}_t + \eta
\end{align*}
$$

(20)
Strategy 3: Two-lag Multi-Regime Model

Figure: Building 4
<table>
<thead>
<tr>
<th>building</th>
<th>strategy 1</th>
<th>strategy 2</th>
<th>strategy 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35693</td>
<td>8130</td>
<td>8043</td>
</tr>
<tr>
<td>2</td>
<td>22737</td>
<td>5323</td>
<td>4568</td>
</tr>
<tr>
<td>3</td>
<td>17371</td>
<td>8825</td>
<td>6491</td>
</tr>
<tr>
<td>4</td>
<td>48977</td>
<td>6007</td>
<td>5032</td>
</tr>
</tbody>
</table>

**Table:** The mean square error on each building using different strategies.
Conclusion

Contributions

- investigate and develop a linear state-space model for the problem of energy consumption prediction.
- elaborate two variant from the original state-space model: multi-regime model and two-lag multi-regime model.
- three strategies reflect the track of improvement in our exploration.
Futur work

- prediction accuracy on the test data is not as good as that on the validation data. Solution: model the exogenous observations by another state space model, and estimate the parameters simultaneously for thermal and exogenous model.
- extend to non-linear model, using extended or unscented Kalman filter and EM algorithm.
- adaptive learning of regime function, without using fixed regime periods.
- more historical exogenous observations can be taken into account in the model.