Electrical Power and Energy Systems 91 (2017) 92–100

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Short term electric load forecasting using an automated system of model choice

Qihong Duan [a,*], Junrong Liu [b], Dengfu Zhao [c]

[a] School of Mathematics and Statistics, Xi'an Jiaotong University, 710049 Xi'an, Shaanxi, People's Republic of China
[b] School of Mathematics, Northwest University, 710069 Xi'an, Shaanxi, People's Republic of China
[c] School of Electrical Engineering, Xi'an Jiaotong University, 710049 Xi'an, Shaanxi, People's Republic of China

A R T I C L E I N F O

Article history:
Received 17 January 2016
Received in revised form 23 January 2017
Accepted 6 March 2017

Keywords:
Load forecasting
Load modeling
Hidden Markov models
Learning systems

A B S T R A C T

The paper studies the choosing mechanism of an energy company which gathers a library of electric load models and at every day chooses the best one for daily prediction. We use a combination of a semi-Markov process and a modified hidden Markov chain to describe the joint curve of loads, the daily best model, and exogenous information, of which temperature is an important factor. By extending the state space of the semi-Markov process, then in a computationally tractable way, the problem is embedded within a hidden Markov chain. Hence we can establish an EM algorithm and an enhancing statistical learning method for estimating parameters and forecasting load. Simulation reveals the range in which the proposed algorithm is applicable. Examples from real world datasets show that the proposed automated system is an alternate method for short term electric load forecasting with loads greater than a few hundreds MW. Supplementary material includes scripts of the proposed system and a guide of the scripts.

1. Introduction

There are many models for electric load forecasting over periods from an hour to several years. The models include different traditional regression methods, and many new non-conventional statistical techniques such as generalized additive models [4] and random forests [2], and many soft computing techniques such as neural networks [10] and support vector machines. [11] reviews important theoretical methods and industry practice on electric load models selected from thousands of papers. Many successful models can attain about a few percent of error when forecasting over a short-term period. Nevertheless, because the storage of electricity wants much extra costs, electric load modeling is still an important issue, and an energy company often hires a team of experts to develop models and to forecast electric loads.

Such experts maintain a library of electric load models. In fact, electric loads as well as the corresponding mechanisms are not the same for different seasons, and we need different models to describe such different mechanisms. Moreover, theoretical and systematic discrepancies between models imply that a model usually prevails in some proper conditions. Choosing the best model from the library is another important and daily work of experts. An expert always carefully examines recent electric loads, temporal and meteorological factors, and other exogenous data, then decides the final choice based on the feature of a model, in accordance with his or her experience.

We will substitute an automated system for the preceding manual process. As the issue is complex, some pretreatment of detrending techniques [11,14] is indispensable. Then the preceding manual process of a choice is a successor to the similar day method [11], which forecasts a future load using the historical days with similar profiles. With a big dataset, we can identify many similar profiles and a type of profile emerges in the investigation. Different types of profiles gradually evolve into different models. The theory of Markov chain provides us with a natural framework for such models [18]. For example, suppose that the chance of a model tomorrow depends on previous models through the last two days. We set the state at a day is determined by models during both the day and the previous day, and then the evolution of states becomes a Markov chain. A careful investigation into the library of models and auxiliary variables may help to establish a concise Markov chain [15].

Denoting the time by t, there is a specific models X_t connecting the forecast of the load L_t to the neighboring $L_{t-1}, L_{t-2}, \ldots$, the temperature T_{t-1}, and other meteorological factors. Based on datasets from several Chinese energy companies, we find that each state of the model X_t depends strongly on the previous state, but the

* Corresponding author.
E-mail address: khtuan@126.com (Q. Duan).

http://dx.doi.org/10.1016/j.ijepes.2017.03.006
0142-0615/

List of variables

L_t, T_t	load and temperature at time t
$X_t = i$	forecasting model i is chose at time t
$f_i(\cdot)$	specific forecasting function in the ith model
N	number of historical data. $N+1$ is the forecasted day
M	number of models in a library. $1, \ldots, M$ is the state space of X_t
δ_t	error. Difference between the true load and its forecasting value
L_{N+1}	load at next day. Our goal is the forecasting of it
$T_{1:N}, L_{1:N}, X_{1:N}$	historical dataset of temperature, load, and model
$T_{1:N}, \delta_{1:N}, X_{1:N}$	Alternative historical dataset, which provides with the same information about $T_{1:N}, L_{1:N}, X_{1:N}$ in our case

durations at a state are not geometric distributed. Hence, a semi-Markov process is a good framework for X_t.

Then the issue can be described as a combination of a semi-Markov process and a hidden Markov chain, and our task be to adapt the theory of hidden Markov chains to the case of an underlying semi-Markov chain. By exploring ideas of interrelated models such as variable duration and length Markov models [3,15], we extend the states of the semi-Markov process X_t and build a Markov chain for the model X_t. The idea is similar to [3,15] where semi-Markov processes with a few states are involved, and we develop a set of techniques based on Coxian distributions and the Akaike information criterion [1] to handle a lot states efficiently.

The rest of the article is organized as follows. In Section 2, we establish our model which is a combination of hidden Markov chains, semi-Markov chains, and discrete Coxian random variables. Some mathematical details of Section 2 are presented in Appendix. By using several simulations, Section 3 explores the range in which the proposed algorithm is applicable. In Section 4, real world examples show that the proposed method is practical for electric load forecasting with loads greater than a few hundreds MW. Section 5 concludes the paper.

2. Method

In general, dependencies of the detrending data L_t, T_t, X_t are intricate. The phenomenon is also pointed out by [19] and references therein. Fortunately, we notice that the error δ_t, which is the difference $L_t - L_t^*$ between the load L_t and the corresponding prediction L_t^* based on a given model X_t as well as neighboring loads and temperatures, is often independent of the history of the data $\delta_s, T_s, X_s, s < t$. For continual long observation, the dataset of $\{\delta_t, T_t, X_t\}$ gives us the same information about $\{L_t, T_t, X_t\}$. Hence, we can focus on the variables $\{\delta_t, T_t, X_t\}$ without loss of generalization. For convenience and to concentrate attention on the issue, we assume that given a model X_t, the conditional distribution of temperature T_t and error δ_t is independent of the history of the data $\delta_s, T_s, X_s, s < t$. Moreover, replacing temperature by another appropriate index suggested by [13,19] as well as other researchers, is not difficult.

Our data include overall a library $\{f_1, \ldots, f_M\}$ of models and four vectors of length N of variables: temperature $T_{1:N} = (T_1, \ldots, T_N)$, load $L_{1:N} = (L_1, \ldots, L_N)$, error $\delta_{1:N} = (\delta_1, \ldots, \delta_N)$, and model $X_{1:N} = (X_1, \ldots, X_N)$. For a model $X_t = i$, the corresponding f_i in the library results in a prediction $L_t^* = f_i(T_t, L_{t-1}, L_{t-2}, \ldots)$ to load L_t and we have error $\delta_t = L_t - L_t^*$. The goal is to: (1) estimate the joint distribution $P(T_{1:N}, \delta_{1:N}, X_{1:N})$ and (2) predict a new load L_{N+1} and its corresponding error δ_{N+1} given a new temperature T_{N+1}, based on the joint distribution $P(T_{1:N}, \delta_{1:N}, X_{1:N})$.

2.1. Generalized hidden Markov chains

The model decomposes the data and the corresponding joint distribution $P(T_{1:N}, \delta_{1:N}, X_{1:N})$ into three components:

(1) An initial distribution of the model, $\pi = (\pi_1, \ldots, \pi_M)$ where $\pi_i = P(X_1 = i)$. If X_1 is a given model i, then $\pi_i = 1$ and $\pi_j = 0$ for $j \neq i$.
(2) A transition probability matrix of the time-homogeneous Markov chain $X_{1:N}, A = (a_{i,j})$ where $a_{i,j} = P(X_{t+1} = j | X_t = i)$.
(3) A set of time-homogeneous multivariate probability density functions (pdf) $G = (g_1(\delta, t), \ldots, g_M(\delta, t))$ where $g_i(\delta, T) = P(\delta_t, T_t | X_t = i)$ is the conditional pdf of δ_t, T_t given $X_t = i$.

Differing from the classical hidden Markov chains [18], where it is difficult and complex task to estimate model parameters π, A, G, we have both signal data $T_{1:N}, \delta_{1:N}$ and underlying model data $X_{1:N}$, and hence we can easily obtain estimators. Typically, we estimate π by empirical frequency of $X_{1:N}$, and estimate an element $a_{i,j}$ of the transition matrix A by the empirical frequency of transitions from state i to j. Moreover, we can estimate every $g_i(\delta, T)$ based on δ_t, T_t such that $X_t = i$. Finally, from the estimated model parameters, constructing the joint distribution $P(T_{1:N}, \delta_{1:N}, X_{1:N})$ is a trivial issue.

$$P(T_{1:N}, \delta_{1:N}, X_{1:N}) = \pi_{X_1} \prod_{t=1}^{N-1} a_{X_t, X_{t+1}} \prod_{s=1}^{N} g_{X_s}(\delta_s, T_s). \quad (1)$$

Our another goal is the forecasting of out-of-sample load L_{N+1}, and we can not simply follow the forward-backward algorithm [17] because we must use a new temperature T_{N+1} as well as the sample data $T_{1:N}, \delta_{1:N}$, and $X_{1:N}$. Nowadays, observatories report the future temperature accurately, and the exogenous information can improve the forecast remarkably. Hence, it is worthwhile establishing a new statistical learning method for a problem involving T_{N+1}. In the following, we will solve the problem step by step.

First, as discussed in Section 2, the dataset of $\{\delta_t, T_t, X_t\}$ gives us the same information about $\{L_t, T_t, X_t\}$ in our case. Hence the prediction to load L_{N+1} can be described by the following formula in terms of our variables δ_t, T_t, X_t.

$$E[L_{N+1} | T_{N+1}, T_{1:N}, L_{1:N}, X_{1:N}] = E[L_{N+1} | T_{N+1}, T_{1:N}, \delta_{1:N}, X_{1:N}]. \quad (2)$$

Moreover, we have $L_{N+1} = L_{N+1}^* + \delta_{N+1}$. And there is a tractable formula

$$L_{N+1}^* = f_i(T_t, L_{t-1}, L_{t-2}, \ldots),$$

given that the out-of-sample model $X_{N+1} = i$ is known as well as $T_{N+1}, T_{1:N}, \delta_{1:N}, X_{1:N}$. Therefore, we compute the expectation in (2) by conditioning on X_{N+1}. That is, to calculate the expectation, we take a weighted average of the conditional expectation $E[L_{N+1} | T_{N+1}, X_{N+1} = i, T_{1:N}, \delta_{1:N}, X_{1:N}]$ given that $X_{N+1} = i$, and every term is weighted by the corresponding probability $P[X_{N+1} = i | T_{N+1}, T_{1:N}, \delta_{1:N}, X_{1:N}]$. That is,

$$\begin{aligned} &E[L_{N+1} | T_{N+1}, T_{1:N}, \delta_{1:N}, X_{1:N}] \\ &= \sum_{i=1}^{M} P(X_{N+1} = i | T_{N+1}, T_{1:N}, \delta_{1:N}, X_{1:N}) \times (f_i(T_t, L_{t-1}, L_{t-2}, \ldots) \\ &\quad + E(\delta_{N+1} | T_{N+1}, X_{N+1} = i, T_{1:N}, \delta_{1:N}, X_{1:N})). \end{aligned} \quad (3)$$