Peak Load Forecasting of Electricity Generating Authority of Thailand by Gaussian Process

Tuchsanai Ploysuwan
Department of Electrical Engineering
Siam University
Bangkok, Thailand
tuchsanai@gmail.com

Pramukpong Atsawathawichok and Prasit Teekaput
Department of Electrical Engineering
Chulalongkorn University
Bangkok, Thailand
g.pramukpong@gmail.com and Prasit.T@chula.ac.th

Abstract—**This paper use Gaussian Process : $\mathcal{GP}$ to present about the peak electricity (Peak load) forecasting for the highest demand of electricity since 2011 to 2012 of "Electricity Generating Authority of Thailand" (EGAT) by use the data since 2000 to 2010 as a the training data. The four important variables are 1) time per month, 2) peak load per month, 3) GDP 4) GNP and present about how to compute the hyper – parameter θ which is the important variable that cause an efficiency forecast. The result of experiment has shown that the process which give few error and has more efficiency than Neural Network (NN).**

Index Terms—**Gaussian Process, Load Forecasting, Peak Electricity Demand, Neural Network**

I. Introduction

In each year the demand of electricity in the country tends to be rapidly increasing, as the development of economy and industry increases. Electric energy is also a key factor for the economic system's propulsion in order not to effect economic system and public security. The demand of electricity forecast is usually use to load forecasting. ([1],[2],[3],[4]). It is very important to plan for the production of electricity system and electricity storage, trade of electricity and specification of electricity rate for the present and for the future. So the "Electricity Generating Authority of Thailand" (EGAT) has collected and propagated peak load and energy load of each month in the yearly report [5] continuously since the year in 2000 to 2012 to use for the plan to build the power plant, transmission line and information for investment of the prediction for the highest demand and the increase of the net of electricity every year.

Peak load and energy load has a relation by non – linear with many variable in Physics and Economic by the way of Physics such as climate, temperature and by the way of economic such as the trend of energy used in the past, population, GNP, GDP, oil price and electricity per unit ([2],[3]). Where in the forecasting is divided in two parts (short– term: 15 minutes to 1 week) and (long – term: month, year or 10 years). Generally the technique for forecasting are the Autoregressive (ARMA), Neural Network (NN) ([3],[4]) and Support Vector Regression (SVR). Each variable is use indifferent ways upon the forecasting. Such as short – term forecasting for physics factor, oil price and electrical price that have affected the estimates, but the long – term forecasting for economic such as GNP and GDP are progressive [3]. However, by the methods that have been shown, if there is only few training data it will cause problems for the local minimum [2] in finding the answer.

Gaussian Process is Stochastic Process. Now it is accepted as an efficiency tool to solve problems which are Regression, classification and Decision. In machine learning can work well and has efficiency even though it has only a few training data and the convergence rate is better than ARMA, NN and SVR by step [7].

This paper presents about how to use Gaussian process to forecast the long – term of the demand of peak load by using data since 2000 to 2010 for a training data set containing such as 1) peak load per month, 2) time (month and year), 3) GDP, 4) GNP and using data since 2011 to 2012 for testing ability of algorithm and will use only time variable (not use GDP and GNP). Furthermore it presents about an optimization to estimate the hyper – parameters (θ) that is also a variable key to increase efficiency of Gaussian process.

II. Gaussian Process ($\mathcal{GP}$)

A. $\mathcal{GP}$ Fundamental

The meaning of Multivariable Gaussian distribution is defined by probability density function: $\mathcal{N}(\mu,\Sigma)$

$$\mathcal{N}(\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\Sigma|}}\exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right) \quad (1)$$

Where $x \sim \mathcal{N}(\mu,\Sigma)$ is a random variable a vector $x \in \mathbb{R}^d$ Mean $\mu = \mathrm{E}[x] \in \mathbb{R}^d$ and Covariance $\Sigma = \mathrm{E}\left[(x-\mu)(x-\mu)'\right] \in \mathbb{R}^{d\times d}$

Gaussian Process ($\mathcal{GP}$) [7] is Probability density function (pdf) that distribute by Gaussian on function which is defined by

$$f(x_i) \sim \mathcal{GP}\left(m(x_i), k(x_i,x_j)\right) \quad (2)$$

$f : x \in \mathbb{R}^d \to \mathbb{R}$ in theory "d" of $\mathcal{GP}$ can has more value until ∞ or $x=\mathbb{R}$ by value of $f(x)$ randomed that contain the mean: $m(x_i)$ and covariance function: $k(x_i,x_j)$ [Sometimes $k(x_i,x_j)$ is called Kernel Function of $f(x)$]

978-1-4799-3174-3/14/$31.00 ©2014 IEEE

$$m(x_i)=\mathrm{E}[f(x_i)]$$
$$k(x_i,x_j)=\mathrm{E}[(f(x_i)-m(x_i))(f(x_j)-m(x_j))] \tag{3}$$

Vector x_i and $x_j \in \mathbb{R}^d$ has value for example $x_i \in \left[x_{i,1},x_{i,2},...,x_{i,\mathrm{d}}\right]^T$ for easy to understand by assume $m(x_i)=0$ and by the way to Kernel function $k(x_i,x_j)$ in (3) has limitation in case of 1) number of training data x_i and $f(x_i)$, 2) sometimes incorrect of information for example $y_i=f(x_i)+\varepsilon_i$ by ε_i is error because measurement cannot estimates the value of E[.] in (3) correctly. So how to solve Kernel function usually solve indirect and stuck in θ for example squared exponential [7].

$$k(x_i,x_j|\theta)=\sigma_{se}^2 \exp\left(-\frac{1}{2}\sum_{k=1}^{d}\frac{(x_{i,k}-x_{j,k})^2}{l_k}\right) \tag{4}$$

By defining the parameter $\theta=(\sigma_{se}^2,l_1,l_2,...,l_d)$ is called Hyper-parameter [parameter in kernel function: $k(x_i,x_j)$]

B. GP Priori

Consider the length of training data system (N) define by value of set X and set of function f that pair with X is

$$X \in \mathbb{R}^{N\times d} = \left\{x_i = \left[x_{i,1},x_{i,2},...,x_{i,\mathrm{d}}\right]^T\right\}_{i=1}^{N}$$
$$f \in \mathbb{R}^{N\times 1} = \left[f(x_1),f(x_2),...,f(x_N)\right]^T \tag{5}$$
$$D=\{X,f\}$$

By D in (5) is called training data set when $\mathcal{GP}$ prior has the length (N) of training data define by

$$p(f|X,\theta)=\mathcal{N}(0,K_{f,f}) \tag{6}$$

Value of covariance matrix $K_{f,f} \in \mathbb{R}^{N\times N}$ can create by equation (4) by $K_{f,f}(i,j)$ is a member of matrix row i column j of kernel matrix: $K_{f,f}$

$$K_{f,f}(i,j)=k(x_i,x_j|\theta) \tag{7}$$

III. How to Model the Peak Load Forecast by $\mathcal{GP}$

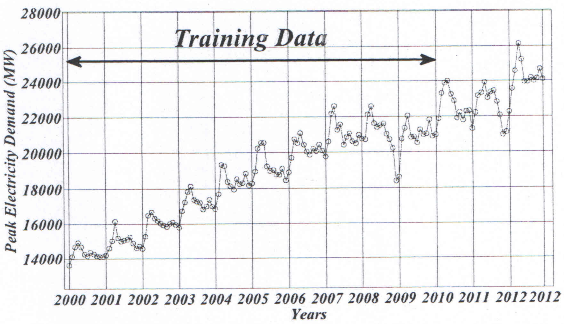

Fig. 1. The demand of peak electricity per month since 2000 to 2012 [5]

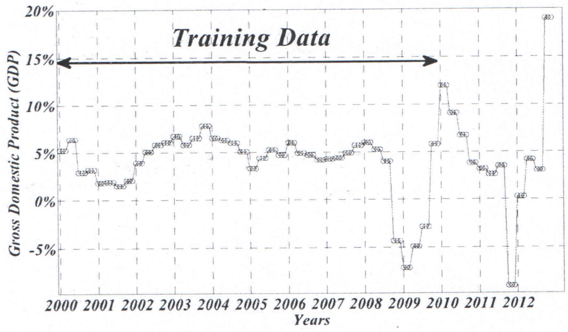

Fig. 2. GDP since 2000 to 2012 [8]

As this paper use peak load data [5] by monthly since 2000 to 2010 so the length of training data is $N=11\times12=132$ that has shown in fig. 1. (by gathering data for 11 years) define $i\in\{1,2,...,N\}$ by t_i is time of peak load data storage for example $i=3$ is the data of March 2000 [5] so $t_i=datenum(2000,3,1)$ $=730546$ by the "datenum" command is how to convert the value [year month time] to time index in MATLAB program so vector $x_i\in\mathbb{R}^3=[t_i,GDP(t_i),GNP(t_i)]^T$ and the peak electricity: $f(x_i)$ that pair with $f: x \in \mathbb{R}^3 \to \mathbb{R}$ by the set of training data: $D=\{X,f\}$ is defined like equation (5) and the $\mathcal{GP}$ Prior can solve by (6),(7) by step

This paper forecast peak load or $f(x_j^*)$ of data since $t_j \in [2011\ 2012]: j \in \{1,2,...,M\}: M = 24$ assume by $x_j^*=(t_j,0,0)$ $X^* = \{x_j^*\}_{j=1}^{M}, f^* = \left[f(x_1^*),f(x_2^*),...,f(x_M^*)\right]$ so the Joint Gaussian pdf of latent f and f_* is

$$p\left(\begin{bmatrix} f \\ f_* \end{bmatrix} \middle| X,X_*,\theta\right)=\mathcal{N}\left(0,\begin{bmatrix} K_{f,f} & K_{f,f_*} \\ K_{f,f_*}^T & K_{f_*,f_*}\end{bmatrix}\right) \tag{8}$$

$K_{f,f}\in\mathbb{R}^{N\times N}=k(X,X|\theta), K_{f,f_*}\in\mathbb{R}^{N\times M}=k(X,X_*|\theta)$ and $K_{f_*,f_*}\in\mathbb{R}^{N\times M}=k(X_*,X_*|\theta)$ reference from appendix A in (8)

$$p(f_*|X,X_*,\theta)=\mathcal{N}(m(X_*|\theta),K(X_*|\theta)) \tag{9}$$

Or $f_*|X,X_*,\theta \sim \mathcal{N}(m(X_*|\theta),K(X_*|\theta))$ so the forecast of peak electricity demand of $\mathcal{GP}$ is

$$f_* \sim \mathcal{GP}(m(X_*|\theta),K(X_*|\theta)) \tag{10}$$

Mean and covariance (Kernel matrix) of $\mathcal{GP}$ in (5) is

$$m(X_*|\theta)= K_{f,f_*}^T K_{f,f}^{-1} f$$
$$K(X_*|\theta) = K_{f_*,f_*} - K_{f,f_*}^T K_{f,f}^{-1} K_{f,f_*} \tag{11}$$

Peak load since 2011 to 2012 is represent by condition mean $f_*\in\mathbb{R}^M=m(X_*|\theta)=\left[m(X_*^1|\theta),m(X_*^1|\theta),...,m(X_*^M|\theta)\right]^T$ and then write the new equation (6) in format of Kernel linear combination

$$f\left(x_j^*\right) \cong m\left(x_j^*|\theta\right) = \sum_{i=1}^{N} \alpha_i k\left(x_i, x_j^*|\theta\right) \quad \text{where } \alpha \in \mathbb{R}^N = K_{f,f}^{-1} f \tag{12}$$

By vector $\alpha=[\alpha_1,\alpha_2,...,\alpha_N]^T$ and $j\in\{1,2,...,M\}$

IV. HOW TO COMPUTE THE HYPER–PARAMETER θ

The Hyper–parameter $\theta=(\sigma_{se}^2,l_1,l_2,...,l_d)$ is a variable key in Kernel function $k(x_i,x_j|\theta)$ in (4) that is the answer of peak load forecasting $f(x_j^*)\cong m(x_j^*|\theta)$ by using Gaussian Process: $\mathcal{GP}$ in (12) so the efficiency to forecast depend on the correctly of solving θ when has infinite training data. By the way, to solving θ use information from training data x_i and $f(x_i)$ by using the assumption of peak electricity from estimation (from "EGAT") [5] has an error from $\varepsilon_i\sim\mathcal{N}(0,\sigma^2)$ so the training data of peak load is modeled to

$$y_i=f(x_i)+\varepsilon_i \tag{13}$$

And give vector $Y=[y_1,y_2,...,y_N]^T$ if assume as $\sigma^2\to 0$ means that the estimating data has high plausibility so we will find the value of marginal likelihood of variable f in (13) can modeled to

$$p\left(Y|X,\theta,\sigma^2\right)=\int p\left(Y|f,X,\theta,\sigma^2\right)p\left(f|X,\theta,\sigma^2\right)df \tag{14}$$

By using $\mathcal{GP}$ Prior in (6) and likelihood function in (13) will get

$$\begin{aligned} p\left(f|X,\theta,\sigma^2\right)&=p\left(f|X,\theta\right)=\mathcal{N}\left(f|0,K_{f,f}\right) \\ p\left(Y|f,X,\theta,\sigma^2\right)&=\prod_{i=1}^{N}\mathcal{N}\left(Y|f(X_i),\sigma^2\right) \end{aligned} \tag{15}$$

Bring (15) replace in (14) and integral by using appendix B

$$p\left(Y|X,\theta,\sigma^2\right)=\mathcal{N}\left(Y|0,K_{f,f}+\sigma^2 I_{N\times N}\right) \tag{16}$$

So the log marginal likelihood of Hyper-Parameters θ is

$$\begin{aligned} L' &= \log p\left(Y|X,\theta,\sigma^2\right) \\ &= -\frac{1}{2}Y^T(K_{f,f}+\sigma^2 I)Y-\frac{1}{2}\log\left|K_{f,f}+\sigma^2 I\right|-\frac{N}{2}\log 2\pi \end{aligned} \tag{17}$$

Give a negative log likelihood of equation (17) can estimate to

$$L(\theta,\sigma^2)=\frac{1}{2}Y^T(K_{f,f}+\sigma^2 I)Y-\frac{1}{2}\log\left|K_{f,f}+\sigma^2 I\right| \tag{18}$$

We estimate θ and σ^2 by using MAP (Maximum a posterior) estimation and by Bayes' theorem that will get

$$\begin{aligned} \left\{\theta,\sigma^2\right\} &= \underset{\theta,\sigma^2}{\arg\max}\, p\left(\theta,\sigma^2|X,Y\right) \\ &\propto \arg\max\log\left(\frac{p\left(Y|X,\theta,\sigma^2\right)p(\theta)p\left(\sigma^2\right)}{p(X,Y)}\right) \\ &\propto \arg\max\log p\left(Y|X,\theta,\sigma^2\right)p(\theta)\,p\left(\sigma^2\right) \end{aligned} \tag{19}$$

$p(X,Y)$ has not effected with maximization in (19) so for easy to understand and be convenient to solve $\{\theta,\sigma^2\}$ the maximization problem will equivalent with

$$\left\{\theta,\sigma^2\right\}=\arg\min L\left(\theta,\sigma^2\right)-\log p(\theta)-\log p\left(\sigma^2\right) \tag{20}$$

So how to solve $\left\{\theta,\sigma^2\right\}$ can solve from numerical method by minimization equation (20) (using "fminsearch" command in MATLAB) by choosing prior probability $\left\{\theta,\sigma^2\right\}$ to Inverse - Gamma priori

$$\begin{aligned} p\left(l_i\right) &= \frac{\beta^\alpha}{\Gamma(\alpha)} l_i^{(\alpha+1)}\exp\left(-\frac{\beta}{l_i}\right) \quad i=1,2 \\ p(\theta) &= p\left(l_1\right)p\left(l_2\right) \end{aligned} \tag{21}$$

Give $\beta=10$, $\alpha=0.5$ use for the initial conditions and priori probability density function of the error estimation $\varepsilon_i\sim\mathcal{N}(0,\sigma^2)$ and assume to Inverse-chi-squared distribution

$$p\left(\sigma^2\right)=\left(\frac{1}{\sigma^2}\right)^{(\alpha-1)}\exp\left(\frac{-\beta}{\sigma^2}\right) \tag{22}$$

Give the initial conditions in (22) to $\beta=0.01$, $\alpha=5$ by step and in this paper we assume $\sigma_{se}^2=1$ for decrease the complexity of minimization for solving the answer in equation (20)

V. THE RESULT OF EXPERIMENT

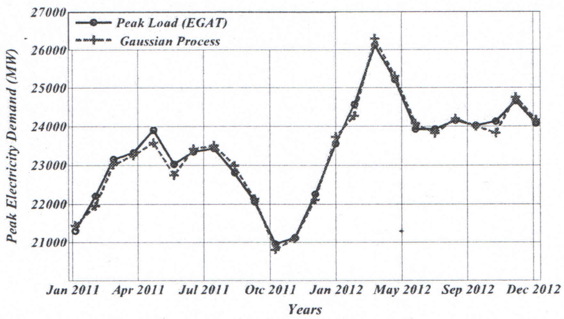

Fig. 3. The result of the demand of peak electricity forecast since 2011 to 2012 by using Gaussian Process

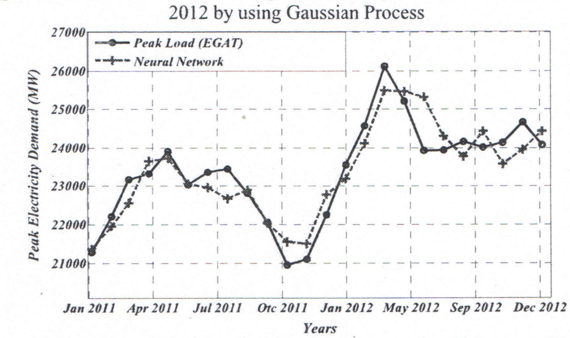

Fig. 4. The result of peak load forecast by using Neural Network [3]

The result of the demand of peak electricity forecast since 2011 to 2012 (total 24 month) by using Gaussian Process: $\mathcal{GP}$ that include the solving of Hyper-Parameters θ from equation (20) that has shown in fig. 3. and the result compare by using Neural Network: NN that has shown in fig. 4. and use the training data since 2000 to 2010 contain by four variables for example time per month, peak load per month, GDP and GNP by step

VI. Conclusion

This paper use Gaussian Process : $\mathcal{GP}$ to present about the peak electricity (Peak load) forecasting for the highest demand of electricity since 2011 to 2012 of "Electricity Generating Authority of Thailand" (EGAT) by using the data since 2000 to 2010 as a the training data. So the simulation result that has shown above is Gaussian Process: $\mathcal{GP}$ that is similar to the real data than Neural Network: NN that increase confident to choose the forecast data of Gaussian Process: $\mathcal{GP}$.

Acknowledgment

Appreciative to working group of iEECON2014 for the opportunity and little space for practicing about research and presenting the academic research and at last thanks for all suggestion from every reviewers

References

[1] M. Chanza, P. Ramjith, and G. Van Harmelen, "Forecasting domestic hourly load profiles using vector regressions," in *Domestic Use of Energy Conference (DUE),* 2013 *Proceedings of the* 21*st*, 2013, pp. 1-5.

[2] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy Load Forecasting Using Empirical Mode Decomposition and Support Vector Regression," *Smart Grid, IEEE Transactions on,* vol. 4, pp. 549-556, 2013.

[3] M. A. Mamun and K. Nagasaka, "Artificial neural networks applied to long-term electricity demand forecasting," in *Hybrid Intelligent Systems,* 2004. *HIS '*04. *Fourth International Conference on*, 2004, pp. 204-209.

[4] Y. Zhang, Q. Zhou, C. Sun, S. Lei, Y. Liu, and Y. Song, "RBF Neural Network and ANFIS-Based Short-Term Load Forecasting Approach in Real-Time Price Environment," *Power Systems, IEEE Transactions on,* vol. 23, pp. 853-858, 2008.

[5] "Annual Report " Electricity Generating Authority ofThailand (EGAT) 2001-2012.

[6] J.-j. Wang, D.-X. Niu, and L. Li, "An ARMA Cooperate with Artificial Neural Network Approach in Short-Term Load Forecasting," in *Natural Computation,* 2009. *ICNC '*09. *Fifth International Conference on*, 2009, pp. 60-64.

[7] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Cambridge, Mass.: MIT Press, 2006.

[8] *Office of the National Economic and Social Development Board Thailand.* Available: http://eng.nesdb.go.th

APPENDIX A.

Marginal Gaussian Probability

$$p(x,y) = \mathcal{N}\left(\begin{bmatrix} a \\ b \end{bmatrix}, \begin{matrix} A & B \\ B^T & C \end{matrix}\right) \rightarrow p(x|y)$$
$$= \mathcal{N}\left(a+BC^{-1}(y-b), A-BC^{-1}B^T\right) \tag{23}$$

APPENDIX B.

Product of Gaussian Probability

$$\mathcal{N}(X|a,A)\mathcal{N}(X|b,B) \rightarrow Z^{-1}\mathcal{N}(X|c,C) \tag{24}$$

By $c=C\left(A^{-1}a+B^{-1}b\right), C=\left(A^{-1}+B^{-1}\right)^{-1}$ and Z^{-1} is a constant