A Short-term Load Forecasting Based on Support Vector Regression

Lu Yu
Southeast University
School of Electrical Engineering
Nanjing, P.R.China, 211189
e-mail: runnnningfish@163.com

Abstract — Load forecasting has become a significant part in national power system strategy management. In this paper, the Support Vector Regression (SVR) for Short-term Load Forecasting (STLF) is presented to predict the primacy of the industry power in the electricity composition. The Support Vector Machine (SVM) is introduced to learn a regression model from training samples with relaxation factors. Our experimental data come from a real-time data acquisition system, which is running for industrial users in a city of Eastern China. As input to the regression model, the feature vector of training samples combines meteorological factors with power system data collected from meters. In order to study the effect of different kernel functions on the accuracy of prediction, this paper respectively tests the linear, polynomial kernel function and Radial Basis Function (RBF). We evaluate the method with two types of predictions, discrete prediction of random samples and continuous prediction of sequential samples. The results indicate that the linear regression model is suitable to forecast with a high fitting degree. However, in the continuous date power prediction, the polynomial kernel function shows preferable prediction ability from the impact of emergencies.

Keywords- Load Forecasting; Support Vector Machine; Machine Learning; Telecommunication; Regression Model

I. INTRODUCTION

Load forecasting plays an important role in power system planning and management. Also, it is the essential basis of economical operation of power systems. Therefore, accurate load forecasting gives support to arrange operation plan, distribute electricity contract and coordinate power generation and transmission facilities.

The load forecasting can be divided into four broad categories by prediction time: Very Short-term Load Forecasting (VSTLF) [1], Short-term Load Forecasting (STLF), Medium-term Load Forecasting (MTLF) and Long-term Load Forecasting (LTLF) [2]. STLF refers to one day to one week, which is used to arrange daily and weekly schedule.

Various STLF methods appeared in the field during last decades. They can be divided into three broad categories: traditional forecasting techniques, modified traditional techniques and artificial intelligence [3]. Commonly, the traditional foresting techniques are shown as: regression [4], Mbamalu and El-Hawary modified multiple regression [5], etc. Examples of modified traditional technique are adaptive demand forecasting [6] and time series methods, such as Autoregressive (AR) Model [6] and Autoregressive Moving-average (ARMA) model [7], etc.

Support Vector Machine (SVM), which is proposed by Vapnik in [8], is a novel machine learning method based on statistical learning theory. Its essence is a classic quadratic programming problem, to achieve structural risk minimization. Advantages of SVM are compared with the Artificial Neural Network (ANN) model in [9], which include fast convergent rate, the only global optimal solution that has solved the problem of local extremum in ANN. SVM is originally used to solve classification problems, showing an attractive classification result. The Support Vector Regression (SVR) is proposed by combining SVM model with regression prediction [10]. An LS-SVM model is also proposed by integrating least squares (LS) algorithm with SVM [11]. It has been shown that SVR is quite effective for STFL problems [12, 13].

In this paper, the Support Vector Regression model for STLF, which is combined with the regression analysis based on the original SVM algorithm, is implemented on a Data Acquisition System (DAS) to forecast the primacy of the industry power in the electricity composition, to direct the arrangement of electricity and distribution of electricity contract.

II. REAL-TIME DAS FOR INDUSTIAL USERS

In order to obtain electricity data of industrial users, this paper designs a real-time data acquisition system of electricity. The system is composed of meters, Data Transmission Units (DTU), General Packet Radio Service (GPRS) network and a server, as shown in Fig .1.

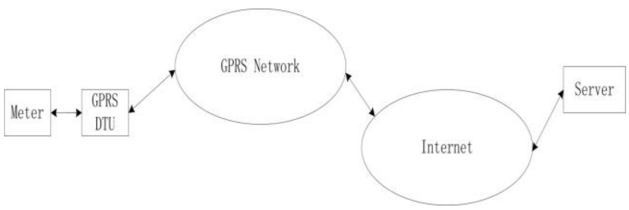

Figure 1. A data acquisition system of electricity.

The electricity data from meters, include: positive active power, reverse reactive power, voltage, current and power factor. These data reflect the quality of the power system and the electricity consumption of industry customers.

GPRS is a mobile data service for the Global System for Mobile (GSM) users. GPRS, which is widely covered in China, has many advantages, such as low cost, convenient installation of GPRS DTUs. GPRS DTUs send electricity data to server electricity data via GPRS.

The server, which contains the function of communication, database and web service, records electricity data from Internet into the database, runs Load Forecasting algorithm, distributes predictions to related users by web pages. The system has been running in a city of Eastern China for more than half a year.

III. SVR MODEL

The SVR model is proposed by combining SVM model with regression prediction, where SVM is introduced to learn a regression model from training samples with relaxation factors.

A. SVM theory

SVM maps the input space into high-dimensional feature space, constructing linear decision function to replace the nonlinear decision function [14]. Given a sample $\{(X_i, y_i)\}_{i=1}^{l}$, $X_i \in R^n$ is an input vector, $y_i \in R, i = 1,2,3\ldots l$ is a output value, l is the number of the training samples. In order to find an optimal hyperplane when the sample is not completely classified, we need to release some conditions. Therefore, a group of non-negative relaxation factors $\xi = (\xi_1, \xi_2, \ldots, \xi_l)^T$ have been proposed. If there are $\omega \in R^n, x \in R^n, b \in R$, satisfying the regression equation $f(x) = (\omega \cdot x) + b$, then we construct the optimization problem about ω , b and $\xi = (\xi_1, \xi_2, \ldots, \xi_l)^T$ as

$$\begin{aligned} &\min_{\omega,b,\xi} \frac{1}{2} \|\omega\|^2 + C\sum_{i=1}^{l} \xi_i \\ &s.t. \quad y_i((\omega \cdot \mathrm{x}_i) + b) + \xi_i \geq 1, \mathrm{i} = 1,2,\ldots,l \end{aligned} \tag{1}$$

where $C > 0$ is a constant, used to control the punishment degree of misclassification, $b \in R$ is one-dimensional offset, $(\omega \cdot x)$ is the inner product of ω and x . By introducing Lagrange operator $\alpha_i \geq 0, i = 1,2,\ldots,l$, we can obtain the linear function form of SVM as

$$f(x) = \sum_{i=1}^{l} \alpha_i x_i^T x_i + b + \xi_i \tag{2}$$

By choosing a proper kernel function instead of the direct inner product， (2) can be deformed as

$$f(x) = \sum_{i=1}^{l} \alpha_i K(x_1, x_2) + b + \xi_i \tag{3}$$

The introduction of kernel functions into SMV allows the algorithm solve nonlinear questions in high-dimensional space with linear decision function.

IV. LEANING OF THE REGRESSION MODEL

When the actual sample is taken into theory model, the model learning is needed. Learning of the regression model includes model analysis, pre-procession of training sample and the introduction of relaxation factor.

A. Model analysis

1) Relevant factors to forecasting

Load forecasting depends on an appropriate mathematical model to predict the consequent load values according to the observed data. The predication is mainly related to following factors [15]:

a) Meteorological components:

Meteorological components are proposed as an input vector in VSTLF and STLF, including temperature, humidity, rainfall, sunlight, etc. Meteorological components have a significant influence on predictions for residential electricity in VSTLF. As the industrial load forecasting belongs to indoor operation, the temperature is chosen as the meteorological component.

b) Typical load components:

Typical load components represent the periodical load change, which has nothing to do with the meteorological factors. These components just indicate the change of current load value compared with before and after the same period or the same time load value.

c) Exceptional factors and social activities

Electrical load has certain randomness, such as holidays, severe weather, mutation of blackouts, the sudden increase of production tasks, etc. These factors above may cause irregular fluctuations on load.

2) Selection of sample input vectors

Combined with the data by the DAS, 10 dimensional features are selected as input vectors based on the analysis of influences on power load. These input vectors are as follow:

a) *Time components*: There are two dimensional time components, one dimension is the identification of work days or rest days (including holidays), the other is to indicate which day of the week, like Sunday to Monday.

b) *Meteorological components*: There are two dimensional meteorological components, respectively corresponding to the highest and lowest temperature in one day.

c) *Electricity parameters*: There are six dimensional electricity parameters of one day: reverse reactive power, A and C phase voltage, A and C phase current and power factor.

B. Pre-processing of training sample

The pre-processing of the training samples y_i is listed as:

1. According to the daily freezing electricity consumption collected by DAS to calculate actual daily electricity consumption e_i , getting n_i by initializing e_i .

2. Arranging the initialized n_i in sequence, if the number of the training sample is l , set the threshold value is $thres = \dfrac{n_{\lfloor \frac{l}{2} \rfloor} + n_{\lfloor \frac{l}{2} \rfloor + 1}}{2}$, to ensure an average classification. Comparing e_i and $thres$, if n_i is larger than the threshold, set the output value y_i to +1, else y_i is -1, thus we obtain the output values y_i .

C. Relaxation factor in SVM

As the SVM algorithm is traditionally used for classification, with the constraint that $y \in \{\pm 1\}$. As in load forecasting problem, the predicted value is the actual electricity consumption, thus we modify the SVM training framework, importing the training sample to regression model with relaxation factors $\xi = (\xi_1, \xi_2, \ldots, \xi_l)^T$ as allowed error of samples.

According to aforementioned pre-procession of samples, after initialized and compared with the threshold, the tag set is classified into two parts: +1 and -1. Then the relaxation factor $\xi_i = \exp(-|e_i - thres|)$. After the pre-procession, we learn our regression model by a modified conventional SVM training framework.

V. EXPERIMENTS

The sample is the real-time electricity data collected from AnYi HongDa Textile Co. ltd., from September 1, 2014 to January 13, 2015.These 132 sets of data includes daily freezing electricity consumption, reverse reactive power, temperature, holidays and parameters of electricity system. After removing the discontinuous data and the zero value acquired from electric meters, 124 sets of data are obtained in total.

A. Results of experiments

1) The discrete prediction of random sample

In order to ensure the universality of the training samples and training for several times, the paper chooses 94 sets of data as the training set which are randomly generated from the 124 sets of data, the other 30 sets of data as the test set to compare different kernel functions.

Absolute error: $E = |predict(i) - true(i)|$

Relative error: $\delta = 100\% \times \dfrac{|predict(i) - true(i)|}{true(i)}$

Mean absolute percentage error:

$$MAPE = 100\% \times \frac{1}{N}\sum_{i=1}^{N}\left|\frac{predict(i) - true(i)}{true(i)}\right| \quad (4)$$

where $predict(i)$ means forecasting value , $true(i)$ presents real value.

In this experiment, 10 groups of training sets and prediction sets are randomly generated. There are 94 samples in training set and 30 samples in test set, polynomial, linear and RBF kernel functions are compared here. In polynomial and lineal kernel function, few outliers (MAPE around 1.0) are removed to obtain MAPEs compared in TABLE I:

TABLE I. MAPE OF THREE KERNEL FUNCTIONS

Kernel / MAPE	Polynomial	Linear	RBF
Average	13.2416%	13.0622%	55.8821%

In TABLE I, polynomial and linear kernel functions show preferable fitting abilities. However, RBF kernel function presents a poor fitting. To display the distribution of errors intuitively, the distribution histograms of errors about various kernel functions are drawn as Fig .2-4:

According to these three distribution histograms, the MAPE of polynomial kernel function and linear kernel function mainly lie in the range of [0, 0.4] and [0, 0.2] respectively. However, the MAPE of polynomial kernel function are distributed from 0 to 1.

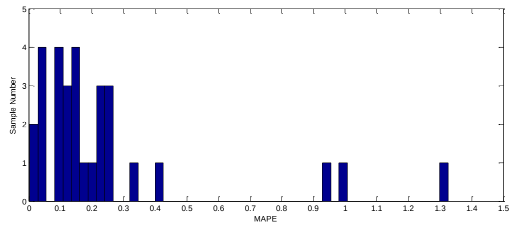

Figure 2. MAPE distribution histograms of polynomial kernel function.

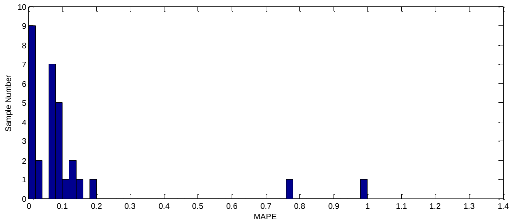

Figure 3. MAPE distribution histograms of lineal kernel function.

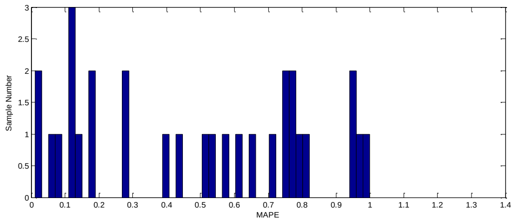

Figure 4. MAPE distribution histograms of RBF kernel function.

2) A continuous prediction of sequential sample

The original 124 samples obtained by day-to-day order can be divided into 4 continuous groups, three of them are taken as training set one as the test set. Fig .5-7 display an example that the first 3 groups are taken as training sets and the fourth group as test set to fit three kinds of kernel functions, from December 15, 2014 to January 13, 2015.In these figures, the solid line presents the predicted values and the dotted line indicates the actual values, taking date as X-axis, load ratio as Y-axis.

The fitting curves indicate that there is a load value mutation with an outlier at the 8th day. Table-3 shows the actual MAPE and the modified MAPE with outliers ignored.

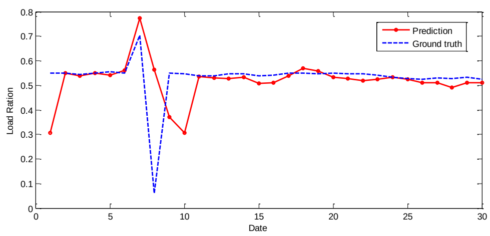

Figure 5. Fitting of polynomial kernel function.

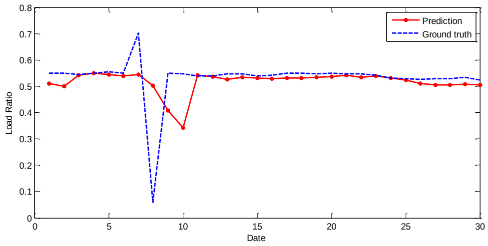

Figure 6. Fitting of linear kernel function.

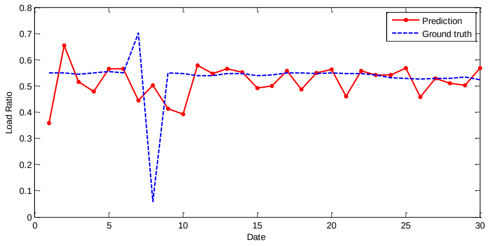

Figure 7. Fitting of RBF kernel function.

TABLE II. MAPE FOR THREE KERNEL FUNCTION

Kernel Function	Polynomial	Linear	RBF
Modified MAPE	6.7909%	5.2819%	9.3126%
Actual MAPE	34.8327%	29.8627%	35.3456%

In TABLE II, polynomial and linear kernel functions display a high accuracy. According to the fitting curves, linear kernel function fits better in a smooth continuous sample and polynomial kernel function fits better at the fluctuate point, the fitting curve of RBF kernel function always deviates from real values.

B. Analysis of Experiments

Based on two types of experiments, we can conclude that the cause of different fitting effects lies in different kernel functions. The concrete impacts of kernel functions can be summarized as follow:

1) *In the discrete prediction of random sample*, the MAPE of polynomial and linear kernel functions are within 15%, as the MAPE of RBF kernel function is over 50%.
2) *In the continuous prediction of sequential sample*, linear kernel functions still maintain perfect accuracy. Especially, polynomial kernel function shows preferable prediction ability in the impact of emergencies.

This observation agrees with the features of kernel functions, linear kernel function is the inner product in the original space. It does not change the dimension of the original space. The complexity of model is increased by more input vectors for polynomial kernel function. The most complicated model of the three is RBF kernel function, which maps the original space projection to the nonlinear space. A typical load forecasting problem fits with models with linear complexity better. However, some emergent factors require higher complexity in the prediction model.

VI. CONCLUSION

The paper presents an SVM classification training framework that imports the training sample to regression model with relaxation factors. This framework obtains satisfying results of load forecast in industrial load, which is significant to direct the arrangement of electricity and distribution of electricity contract. Two types of experiments respected to three kernel functions are test in the paper: the discrete prediction of random sample and the continuous prediction of sequential sample. The results indicate that the linear regression model is preferable to forecast with a high fitting degree, it is more suitable to forecast for STLF. As in the continuous date power prediction, the polynomial kernel function shows improved prediction ability in the impact of emergencies. For future work, we consider to extend our proposed framework into an incremental SVR manner, in which each new sample of electricity parameters can be learned instantly. Hence, the framework can work entirely online for STLF prediction.

REFERENCES

[1] Qia Ding; Jiangang Lu; Huaiqing Liao, "A practical super short term load forecast method and its implementations,"Power Systems Conference and Exposition, 2004. IEEE PES, vol.1, pp.483-485, Oct. 2004

[2] De Silva, D.; Xinghuo Yu; Alahakoon, D.; Holmes, G., "A Data Mining Framework for Electricity Consumption Analysis From Meter Data," Industrial Informatics, IEEE Transactions on, vol.7, no.3, pp.399-407, Aug. 2011

[3] Singh, A.K.; Ibraheem, I.; Khatoon, S.; Muazzam, M.; Chaturvedi, D.K., "Load forecasting techniques and methodologies: A review," Power, Control and Embedded Systems (ICPCES), 2012 2nd International Conference on, pp.1-10, Dec. 2012

[4] Haida, T.; Muto, Shoichi, "Regression based peak load forecasting using a transformation technique," Power Systems, IEEE Transactions on, vol.9, no.4, pp.1788-1794, Nov 1994

[5] Mbamalu, G.A.N.; El-Hawary, M.E., "Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation," Power Systems, IEEE Transactions on , vol.8, no.1, pp.343-348, Feb 1993

[6] Huang, S.R., "Short-term load forecasting using threshold autoregressive models," Generation, Transmission and Distribution, IEE Proceedings- , vol.144, no.5, pp.477-481, Sep 1997

[7] Wang Jian-jun; Niu Dong-Xiao; Li Li, "An ARMA Cooperate with Artificial Neural Network Approach in Short-Term Load Forecasting," Natural Computation, 2009. ICNC '09. Fifth International Conference on, vol.1, pp.60-64, Aug. 2009

[8] Vapnik, V.N., "An overview of statistical learning theory," Neural Networks, IEEE Trans. on, vol.10, no.5, pp.988-999, Sep 1999

[9] Ying-Chun Guo; Dong-xiao Niu; Yan-Xu Chen, "Support Vector Machine Model in Electricity Load Forecasting," Machine Learning and Cybernetics, 2006 International Conference on , pp.2892-2896, Aug. 2006

[10] Kwok, J.T.-Y., "Support vector mixture for classification and regression problems," Pattern Recognition. Fourteenth International Conference on, vol.1, pp.255-258, Aug 1998

[11] Liu, K.; Subbarayan, S.; Shoults, R.R.; Manry, M.T.; Kwan, C.; Lewis, F.L.; Naccarino, J., "Comparison of very short-term load forecasting techniques,"Power Systems, IEEE Transactions on , vol.11, no.2, pp.877-882, May 1996

[12] Ribeiro, B., "Support vector machines for quality monitoring in a plastic injection molding process," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol.35, no.3, pp.401-410, Aug. 2005

[13] Rong Gao; Liyuan Zhang; Xiaohua Liu, "Short-term load forecasting based on least square support vector machine combined with fuzzy control," Intelligent Control and Automation (WCICA), 2012 10th World Congress on, pp.1048-1051, 6-8 July 2012

[14] Elattar, E.E.; Goulermas, J.; Wu, Q.H., "Electric Load Forecasting Based on Locally Weighted Support Vector Regression," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol.40, no.4, pp.438-447, July 2010

[15] Chen Genyong; Shi Jingtian, "Study on the methodology of short-term load forecasting considering the accumulation effect of temperature," Sustainable Power Generation and Supply, 2009. SUPERGEN '09. International Conference on, pp.1-4, April 2009