Expert Systems with Applications 37 (2010) 2388–2394

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A hybrid-forecasting model based on Gaussian support vector machine and chaotic particle swarm optimization

Qi Wu *

School of Mechanical Engineering, Southeast University, Nanjing, Jiangsu 210096, China
Key Laboratory of Measurement and Control of CSE (School of Automation, Southeast University), Ministry of Education, Nanjing, Jiangsu 210096, China

ARTICLE INFO

Keywords:
Support vector machine
Particle swarm optimization
Embedded
Chaotic mapping
Load forecasting

ABSTRACT

Load forecasting is an important subject for power distribution systems and has been studied from different points of view. This paper aims at the Gaussian noise parts of load series the standard ν-support vector regression machine with ε-insensitive loss function that cannot deal with it effectively. The relation between Gaussian noises and loss function is built up. On this basis, a new ν-support vector machine (ν-SVM) with the Gaussian loss function technique named by g-SVM is proposed. To seek the optimal unknown parameters of g-SVM, a chaotic particle swarm optimization is also proposed. And then, a hybrid-load-forecasting model based on g-SVM and embedded chaotic particle swarm optimization (ECPSO) is put forward. The results of application of load forecasting indicate that the hybrid model is effective and feasible.

1. Introduction

Precise short-term load forecasting (STLF) is a basic requirement for the power system. As a very important task for power system operation, STLF helps the electric utility to make important decisions including unit commitment, load switching, etc. In addition, precise load forecasting improves the security of the power system. The research approaches of short-term load forecasting can mainly be divided into two categories: statistical method and artificial intelligence method. In statistical method, an equation can be obtained showing the relationship between load and its relative factors after training the historical data, while artificial intelligence method tries to imitate human being's way of thinking and reasoning in forecasting the future load.

The statistical category includes multiple linear regression (Amjady, 2001; Papalexopoulos & Hesterberg, 1990), stochastic time series (Christianse, 1971), general exponential smoothing, state space, etc. Usually, statistical method can predict the linear load series very well, but it lacks the ability to analyze the nonlinear character of load series due to the inflexibility of its structure. Expert system (Dash, Liew, Rahman, & Ramakrishna, 1995), artificial neural network (ANN) (Chiu, Kao, & Cook, 1997; Xiao, Ye, Zhong, & Sun, 2009) and fuzzy inference (Ying & Pan, 2008) belong to the artificial intelligence category. Expert system tries to get the knowledge of experienced operators and express it in an "if … then" rule, but the difficulty is sometimes the expert's knowledge is intuitive and could not easily be expressed. Artificial neural network does not need the expression of the human experience. It aims to establish a network between the input data set and the observed output data set. It is good at dealing with the nonlinear relationship between the load and its relative factors, but the shortcoming lies in over-fitting and long training time. Fuzzy inference is an extension of expert system. It constructs an optimal structure of the simplified fuzzy inference, which minimizes model errors and the number of the membership functions to grasp nonlinear behavior of short-term loads. However, it still needs the experts' experience to generate the fuzzy rules. Generally, artificial intelligence methods are flexible in finding the relationship between load and its relative factors, especially for the anomalous load forecasting.

Most of the STLF methods hypothesize a regression function (or a network structure, e.g. in ANN) to represent the relationship between the input and the output variables. How to hypothesize the function or the network is a major difficulty because it needs detailed transcendental knowledge of the problem. If the regression form or the network structure is improperly selected, the prediction result would be unsatisfactory. Moreover, it is always a difficulty to select the input variables. Too many or too few input variables would decrease the accuracy of prediction expert system and fuzzy inference do not need to hypothesize the input–output relationship, but it is even more difficult to transform the experts' experience to a rule database. Unlike the statistical models, this NN is a data-driven and nonparametric weak model. Thus, the NN performs well in the problem of load forecasting when the sample data are sufficient. Nevertheless, the available pre-existing load series in companies are often finite. Under this condition, the approximation

* Address: School of Mechanical Engineering, Southeast University, Nanjing, Jiangsu 210096, China. Tel.: +86 25 51166581; fax: +86 25 511665260.
E-mail address: wuqi7812@163.com

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.07.057

ability and generalization performance of the NN are poor. To overcome this disadvantage, a new approach should be explored.

Recently, support vector machine (SVM) (Vapnik, 1995), which is a very promising statistical-learning method, has also been applied to STLF and has shown good result. SVM is firmly grounded in the framework of statistical-learning theory and Vapnik–Chervonenkis theory (VC), which has been developed over the last three decades by Vapnik and Chervonenkis (1974) and Vapnik (1982). Generally speaking, SVM is to minimize the structural risk instead of the usual empirical risk by minimizing an upper bound of the generalization error, and it obtains an excellent generalization performance. Moreover, SVM is especially suitable for solving problems of small sample size and has already been used for classification (Akay, 2009; Chandaka, Chatterjee, & Munshi, 2009; Lee & Lee, 2006; Wu, Liu, Xiong, & Liu, 2009), regression and time series prediction (Tang, Tang, & Sheng, 2009; Wu, 2009, in press; Wu, Yan, & Yang, 2008; Wu, Yan, & Wang, 2009). SVM is to map the input data into a higher dimensional feature space through a nonlinear mapping, and then a linear regression problem is obtained and solved in this feature space (Hu & Song, 2004; Ikeda & Aoishi, 2005; Xiao, Rao, Cecchi, & Kaplana, 2008; Yao & Yu, 2006).

However, the standard SVM encounters some difficulty in real application. Some improved SVMs have been put forward to solve the concrete problem (Wu, 2009, in press; Wu & Yan, 2009a, 2009b; Wu et al., 2008; Wu, Liu, et al., 2009; Wu, Yan, et al., 2009). The standard v-SVM adopting ε-insensitive loss function has good generalization capability in some applications (Wu, 2009, in press; Wu et al., 2008). But it is difficult to deal with the normal distribution noise parts of series. Therefore, the main contribution of this paper can be summarized as follows:

(a) A new version of SVM called SVM with Gaussian loss function (g-SVM) is proposed to approximate load series with normal distribution noise. Compared with standard SVM, the proposed SVM can penalize the Gaussian noise parts of load series effectively.
(b) A new version of PSO called embedded chaotic particle swarm optimization (ECPSO) is also proposed for parameters selection of g-SVM. ECPSO can augment diversity of particles by means of chaotic mapping and enhance the searching ergodicity.
(c) A new hybrid-forecasting method composed of g-SVM and ECPSO is proposed for the STLF. The hybrid-forecasting method can find better solutions in the solution space of the training phase than standard SVM and ARMA.

This paper is organized as follows. The g-SVM is described in Section 2. Section 3 provides a new PSO called embedded chaotic PSO (ECPSO) to obtain the optimal parameters of g-SVM. And then gives the hybrid model based on the g-SVM and ECPSO. In Section 4, g-SVM is used to learn the relationships among all influencing factors and loads. The suitability of the proposed approach is illustrated through an application to real load forecasting from the Jiangsu Electricity Distribution Corporation in china, and then g-SVM is compared with the standard v-SVM and ARMA. Section 5 draws some conclusions.

2. v-support vector machine with Gaussian loss function

2.1. Standard v-SVM model

Suppose training sample set $T = \{(x_i, y_i)\}_{i=1}^{l}$, where $x_i \in R^d$, $y_i \in R$. ε-insensitive loss function can be described as follows:

$$c(x_i, y_i, f(x_i)) = |y_i - f(x_i)|_\varepsilon \tag{1}$$

where $|y_i - f(x_i)|_\varepsilon = \max\{0, |y_i - f(x_i)| - \varepsilon\}$, ε is a given real number.

The standard v-SVM with ε-insensitive loss function can be described as follows:

$$\min_{w,b,\xi^{(*)},\varepsilon} \quad \tau(w, \xi^{(*)}, \varepsilon) = \frac{1}{2}\|w\|^2 + C \cdot \left(v \cdot \varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*) \right)$$
$$\text{s.t.} \quad \begin{cases} (w \cdot x_i + b) - y_i \le \varepsilon + \xi_i \\ y_i - (w \cdot x_i + b) \le \varepsilon + \xi_i^* \\ \xi^{(*)} \ge 0, \varepsilon \ge 0 \end{cases} \tag{2}$$

where w is a column vector with d dimension, $C > 0$ is a penalty factor, $\xi_i^{(*)}(i = 1, \ldots, l)$ are slack variables and $v \in (0, 1]$ is an adjustable regularization parameter, ε is also an adjustable tube' magnitude parameter. Parameter ε appears as the variable of optimal problem, its value is given by the final solution.

2.2. g-SVM Model

The loss function of v-SVM has a significant effect on the generalization capability of the algorithm. Thus, one might consider which loss function should be used. Ideally, it should consist of a simple structure to avoid difficult optimization problems, and it should be suitable for the data. Meanwhile, the noise affects the samples even after denoising. A set of training sample is generated by a function plus additive noise

$$y_i = f(x_i) + \xi_i \tag{3}$$

The likelihood of an estimate $F_f = \{(x_i, f(x_i, w)) | i = 1, 2, \ldots, l\}$ based on the training sample set is

$$P(F_f|F) = \prod_{i=1}^{l} P(f(x_i, w)|(x_i, y_i)) = \prod_{i=1}^{l} P(f(x_i, w)|y_i)$$
$$= \prod_{i=1}^{l} p(y_i - f(x_i, w)) = \prod_{i=1}^{l} p(\xi_i) \tag{4}$$

where $p(\xi_i)$ is the noise density. The appropriate loss function can maximize the likelihood, it is equivalent to maximize $\log P(F_f|F)$

$$\log P(F_f|F) = \sum_{i=1}^{l} \log p(y_i - f(x_i, w)) \tag{5}$$

Thus, the appropriate loss function is

$$c(x_i, y_i, f(x_i, w)) = -\log p(y_i - f(x_i, w)) = -\log p(\xi_i) \tag{6}$$

The abnormal condition can be detected by comparing the similarity measure, which is defined by

$$\log P(F_f|F) = \sum_{i=1}^{l} \log p(\xi_i) = -\sum_{i=1}^{l} c(x_i, y_i, f(x_i, w)) \tag{7}$$

The importance of a new similarity measure is twofold: On the one hand, it presents a principle that enables the construction of a loss function. Once the noise density of the system is defined, its related loss function can be obtained according to (7).

In real-world applications, the standard Gaussian density model $N(0, 1)$ is commonly used to describe noise. Hence, the Gaussian density model and its loss function are employed here as outlined in (8) and (9)

$$p(y_i - f(x_i, w)) = p(\xi_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi_i^2\right) \tag{8}$$

$$c(x_i, y_i, f(x_i, w)) = \frac{1}{2}(y_i - f(x_i, w)) = \frac{1}{2}\xi_i^2 \tag{9}$$

On the other hand, the result provides a theoretical standard by which it can determine whether two signals are generated in the same condition or not. Under normal conditions, the process demonstrates unique characteristics that are reflected in its related signals. The signal patterns in abnormal conditions exhibit