Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian, 15-17 July, 2012

RANDOM FOREST BASED ENSEMBLE SYSTEM FOR SHORT TERM LOAD FORECASTING

YING-YING CHENG, PATRICK P.K CHAN, ZHI-WEI QIU

Machine Learning and Cybernetics Research Center, School of Computer Science and Engineering, South China University of Technology, 510006, Guangzhou, China
EMAIL: yychengcs@163.com, patrickchan@ieee.org

Abstract:

The short term load forecasting plays an essential role in the operation of electric power systems. Plenty of features involved in the forecasting cause a complex system and the long training time. The curse of dimensionality also downgrades the generalization capability of the predictor. This paper applies the random forest based ensemble system to load forecasting application. Rather than selecting a subset of features, which may cause the information lost, all features are considered in the proposed method. Different feature sets are used to construct regression systems and the average method is used as a fusion. The performance of the proposed model is compared with another existing method based on mutual information feature selection using real load datasets in New York and PJM. Experimental results show our method achieves a better result in term of higher accuracy.

Keywords:

Random forest; ensemble; Feature selection; Short-term load forecasting (STLF)

1. Introduction

The load forecasting is always an important issue in the power system planning and operation [1, 2]. In particular, the short-term load forecasting (STLF) is crucial in unit commitment and maintenance, power interchange and task scheduling of both power generation and distribution facilities [3]. Economically, the accurate load forecasting lower the operating cost which contributes to significant savings in electric power companies [4]. However, the load forecasting is a challengeable problem due to its complex and nonlinear relationship with many factors, for example, meteorological conditions, special events and seasonal effects (daily and weekly cycles, calendar holidays).

According to studies of the load forecasting [5-9], the load forecasting can be classified into four categories in terms of lead time:

1) Long-term load forecasting with the lead time of more than one year
2) Mid-term load forecasting with the lead time of 1 week to 1 year
3) Short-term load forecasting with the lead time of 1-168 h
4) Very short-term load forecasting with the lead time shorter than 1 day

During last few decades, many research have been devoted to cope with the techniques of STLF such as conventional statistical methods [10-13], ANN based models [14-15], fuzzy methods [16-17] and hybrid models [17-18].

The traditional statistical methods including linear or multiple regression [10], autoregressive moving average exogenous variable (ARMAX) [11], exponential smoothing models [12] and Kalman filtering technology [13] have been applied to STLF. The disadvantages of these statistical methods are that they may not represent the nonlinear characteristics of complex loads and consume much more time when the number of variables is increased. Artificial neural networks (ANN) have been also widely used in the STLF. For example, a neural network which consists of two self organizing maps (SOMs) is proposed in [14]. Many studies apply RBF neural network because of its extensive learning and high computing speed [15]. These methods increase the forecasting accuracy because the models represent nonlinear relations between loads and its influencing factors. However, there are still some problems for ANN, such as slow convergence in training and the need for manually determining the structure and parameters. Fuzzy theory is usually combined with other models to form a hybrid system. Neuro-fuzzy approach has been applied successively in a price sensitive environment [16]. A fuzzy expert system which uses fuzzy set theory to model imprecision in the load and temperature models is proposed in [17].

A major problem in load forecasting is how to select the relevant input features without deteriorating

978-1-4673-1487-9/12/$31.00 ©2012 IEEE

discriminative capability of the predictor. There are a huge amount of features for the load forecasting, for example, the weather conditions, the fuel price, the Consumer Price Index (CPI) and etc. Moreover, the historical loads may be useful for the prediction. Feature selection methods [3] have been applied to load forecasting problem in recent years. These methods can be categorized into three types: filter, wrapper, and embedded methods [19-21]. However, as only the most relevant features are remained, the information provided by the removed features is lost. Aiming at this problem, this paper applies ensemble system based on random forest technique to forecast the electricity load. All features are used in the model. Base predictors are trained by using a dataset with different feature subsets randomly chosen from the original feature set. Finally, the decisions are combined by average fusion method. Information provided by all features are used in our model to avoid any information loss.

This paper is structured as follows. In Section 2, the literature review on feature selection technique for load forecasting is introduced. The proposed model is described in section 3. Section 4 provides the experimental results of the proposed method on short-term load forecasting datasets. Finally conclusions are drawn in Section 5.

2. Literature Review

Mutual information is the most common feature selection in the lead forecasting. The information found commonly in two random variables X and Y with a joint probability distribution P(X, Y) is defined as the mutual information MI(X, Y) between the two variables. MI(X, Y) measures the reduced uncertainty of X when Y is observed.

$$MI(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}P(X_i,Y_j)\log_2\left(\frac{P(X_i,Y_j)}{P(X_i)P(Y_j)}\right) \quad (1)$$

In the load forecasting, the candidate set of input features X_1, X_2, …, X_n are given, Y is the target load variable. The candidate feature X_m with larger mutual information MI(X_m, Y) will be added into the input feature set. Hence, we can rank the candidate input features based on their mutual information with the target variable.

There two feature selection strategies based on the mutual information theory: one-stage MI feature selection [22] and two-stage MI feature selection [26]. In the one-stage feature selection technique, the most relevant features in term of the highest mutual information with the target load are selected. However, it does not consider the redundancy of candidate features. Hence, in two-stage method, an additional stage which removes the redundant is included in two-stage MI feature selection. However, in the feature selection, it has been recognized that the combination of the m best features does not necessarily lead to good classification performance [27]. Moreover, the information provided by removed features is loss. Therefore, random forest method is applied in this paper to use the information of all features.

3. Random forest combined with ensemble system

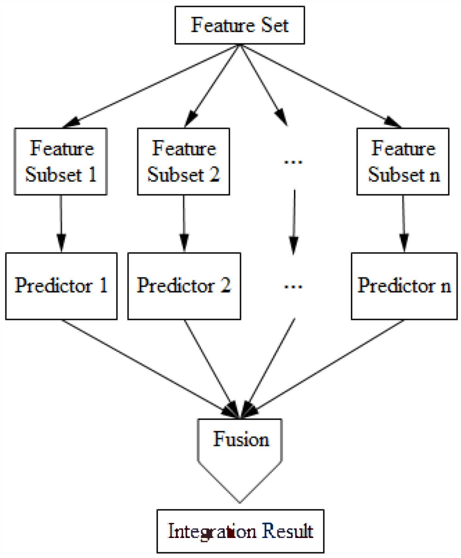

Figure 1. The testing process of the proposed model

Random Forest [25] is a kind of ensemble methods that use randomization to produce a diverse pool of individual regression systems [23]. Each base system is built from a random set of features. Such random feature selection promotes the diversity of systems, and it improves classification performance at the end. A random forest can be built for example by randomly sampling a feature subset for each system, and/or by randomly sampling a training data subset for each system [23]. The final prediction is combined by the majority vote. The random forest achieves a satisfying and robust performance on different applications [24].

This paper applies random forest theory combined with ensemble system to handle the load forecasting. The advantage of using ensemble system is demonstrated in problems that involve a large set of classes and complex input features such as load forecasting.

Figure 1 shows the architecture of the model. During the training, a dataset of load is given. The n preivous consecutive load values are used to form the input features. It means n+1th value is predicted by using its previous n values. m% of full feature set are selected randomly to form a training set for each of L number of base regression system. The final result of each system is combined by using simple average as a fusion method. As the ensemble system is integrated by regression systems which are trained by different feature subsets, the ensemble is expected to have a good performance since different feature subset represents the partial knowledge on the problem.

4. Experimental Study

In this section, the performance of random forest combined with ensemble system using linear function is evaluated experimentally. The method is compared with generalized Regression Neural Network (GRNN) and Back Propagation Neural Network (BPNN) using feature selection models. One-stage and two-stage MI feature selections are applied to mutual information based GRNN and BPNN separately. The real load dataset NYISO [28] and PJM [29] are used. NYISO is the load demand for New York in 2009 and PJM is the electricity market load data in 2011. The load dataset is collected hourly for both datasets. Each experiment is executed 10 times independently. 70% and 30% of dataset are randomly selected as the training and testing set.

The performance of the predictor The relative error at each testing sample is defined as follows:

$$Error = [\frac{P_{predicted}(i) - P_{actual}(i)}{P_{actual}}] \times 100\% \tag{4}$$

in which $P_{actual}(i)$ is the actual load, and $P_{predicted}(i)$ is the forecasting load at time i. Also, the principal statistics used to evaluate the performance of the model, mean absolute percentage error (MAPE), is defined as follows:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N} |\frac{P_{predicted}(i) - P_{actual}(i)}{P_{actual}(i)}| \times 100\% \tag{5}$$

which N is the time interval of averaging.

The parameter setting for random forest model is discussed in Section 4.1 and the experimental comparison result is given in Section 4.2.

4.1 Parameter Selection

For the random forest model in load foresting, three parameters should be decided: the input feature period, the randomly selected feature ratio and the number of predictors. NYISO dataset is used in this section.

Experimental results with different parameter settings are shown in table 1. Four kinds of input feature (i.e. 1 week, 1 month, 3 months and 6 months) periods are considered for this experiment indicated in the first column of Table 1. The second column shows three feature selection ratios (20%, 40% and 60%) for each feature period. The number of predictors (i.e. 200, 400 and 600) is illustrated in the first row. The value in the cell is the average MAPE of the ten independent runs. The bad performance is achieved when the feature period is 6 months. It shows that considering 6 months period of time before is not helpful to predict the load. The performances of other three feature periods are similar. For the feature subset selection ratio, the prediction result using 60% has the lowest MAPE comparing with 20% and 40%. It may because 20% or 40% of full feature set do not provide enough information for the base predictors. The difference between using different the number of base system in the ensemble system is not significant. Using 600 base systems is better than using 200 and 400 in some situations. The random forest trained using 600 classifiers and feature subset randomly chosen 60% from previous 1 month features is best (1.15% MAPE)..

TABLE 1. EXPERIMENTAL RESULTS OF RANDOM FOREST WITH DIFFERENT PARAMETER SETTINGS

Feature period	Base System No. / Selection Ratio	200	400	600
1 week	20%	2.82%	2.78%	2.87%
	40%	1.65%	1.65%	1.67%
	60%	1.19%	1.19%	1.19%
1 month	20%	2.23%	2.38%	2.35%
	40%	1.50%	1.49%	1.47%
	60%	1.16%	1.16%	1.15%
3 months	20%	2.26%	2.31%	2.32%
	40%	1.68%	1.71%	1.70%
	60%	1.60%	1.57%	1.58%
6 months	20%	2.00%	1.99%	1.95%
	40%	4954%	4007%	7346%
	60%	3566%	62830%	14553%

4.2 Model Comparison

The models are compared in this section by using NYISO and PJM. The parameters of the models are selected by using cross-validation. The forecasting results obtained using the proposed method and other feature selection models are shown in table 2 and 3 for NYISO and PJM respectively. The average, variance, minimum and maximum of MAPE are shown in the 3rd, 4th and the 5th rows.

From table 2, the proposed method performs the best among other four models in term of the MAPE (0.83%). Moreover, the proposed method has smallest variance, which is $3.72 * 10^{-8}$, and range (Max – Min), which is

0.06%) comparing with other methods. It means the performance of the method is stable. Although the minimum value of MAPE of the proposed model is 0.23% larger than the one of the one-stage MI based BPNN, the average MAPE of one-stage MI based BPNN is higher than our method.

The result on PJM dataset is shown in table 3. Similar to the result on NYISO shown in table 2, the average, variance and range of MAPE of the proposed method is 1.09%, $5*10^{-8}$ and 0.08% respectively, which are smaller than the other four forecasting models based on MI. It shows that the proposed random forest model performs consistently well on two dataset.

TABLE 2. FORECASTING RESULTS USING THE NYISO DATA

	Random Forest	MI(one stage)+ GRNN	MI(one stage)+ BPNN	MI(two stages)+ GRNN	MI(two stages)+ BPNN
MAPE	0.83%	2.17%	1.03%	2.41%	1.55%
VAR($*10^{-8}$)	3.7210	135.27	991.79	63.707	586.45
MIN	0.80%	2.04%	0.57%	2.30%	1.23%
MAX	0.86%	2.39%	1.50%	2.53%	1.86%

TABLE 3. FORECASTING RESULTS USING THE PJM DATA

	Random Forest	MI(one stage)+ GRNN	MI(one stage)+ BPNN	MI(two stages)+ GRNN	MI(two stages)+ BPNN
MAPE	1.09%	2.60%	1.27%	2.95%	2.16%
VAR($*10^{-8}$)	5.0024	19.478	627.44	484.59	10203
MIN	1.04%	2.51%	0.94%	2.72%	1.14%
MAX	1.12%	2.69%	1.66%	3.48%	4.54%

5 Conclusion

The short-term load forecasting plays a major role under electricity markets. However, due to the huge number of features, the load forecasting is a complex problem. In this paper, rather than selecting the most relevant feature subset, all features are used in a random forest based ensemble system. The random forest based method has been compared with two of the MI based feature selection techniques with MLPNN and GRNN. The experimental comparisons show the random forest based method outperforms than other methods in term of accuracy and stability.

Acknowledgements

This work is supported by National Natural Science Foundation of China (61003171 and 61003172), the Fundamental Research Funds for the Central Universities 2011ZM0066 and a Program for New Century Excellent Talents in University (No. NCET-11-0162).

References

[1]. Gross G, Galiana F, Short-term load forecasting. Proc IEEE 75:1558-1573, 1987.

[2]. Sadownik R, Barbosa E, Short-term forecasting of industrial electricity consumption in Brazil. Int J Forecast 18:215-224, 1999.

[3]. Sheikhan M, Mohammadi N, Neural-based electricity load forecasting using ybrid of GA and ACO for feature selection, Neural Comput & Applic, DOI 10.1007/s00521-011-0599-1.

[4]. Afshin M, Sadeghian A, PCA-based least squares support vector machines in week-ahead load forecasting. In: Proceedings of IEEE industrial and commercial power systems technical conference, pp 1-6, 2007.

[5]. Kothari DP, Nagrath IJ, Modern power system analysis, 3rd edn. Tata McGraw Hill, New Delhi, 2003.

[6]. Feinberg E, Genethliou D, Load forecasting. In: Chow JH, Wu FF, Momoh JJ (eds) Applied mathematics for restructured electric power systems: optimization, control and computational intelligence, power electronics and power systems. Springer, US, pp 269–285, 2005.

[7]. Gonzalez-Romera E, Jaramillo-Moran MA, Carmona-Fernandez D, Monthly electric energy demand forecasting based on trend extraction. IEEE Trans Power Syst 21:1946–1953, 2006.

[8]. Kyriakides E, Polycarpou M, Short term electric load forecasting: a tutorial. In: Chen K, Wang L Trends in neural computation, studies in computational intelligence. Springer, Berlin, pp 391–418, 2007.

[9]. Hahn H, Meyer-Nieberg S, Pickel S, Electric load forecasting methods: tools for decision making. Eur J Oper Res 199:902–907, 2009.

[10]. Papalekopulos A. D, Hesterberg T. C, A regression-based approach to short-term system load forecasting, IEEE Trans. Power Syst., vol.5, no. 4, pp. 1535-1547, Nov. 1990.

[11]. Huang C. M, Huang C. J, Wang M. L, A particle swarm optimization to identifying the ARMAX model for short-term load forecasting, IEEE Trans. Power Syst., vol. 20, no. 2, pp. 1126–1133, May 2005.

[12]. Christianse WR, Short term load forecasting using general exponentials smoothing. IEEE Trans Power Apparatus Syst 90:900–911, 1971.

[13]. Brown RG, Introduction to random signal analysis and Kalman filtering. Wiley, New York, 1983.

[14]. Carpinteiro OAS, Reis AJR, da Silva APA, A hierarchical neural model in short-term load

forecasting. Appl Soft Comput 4:405–412, 2004.

[15]. Singh D and Singh S. P, Self selecting neural network for short-term load forecasting, J. Elect. Power Compon. Syst., vol. 29, pp. 117–130, 2001.

[16]. Khotanzad A, Zhou, and Elragal H, A neuro-fuzzy approach to short-term load forecasting in a price sensitive environment, IEEE Trans. Power Syst., vol. 17, no. 4, pp. 1273–1282, Nov. 2002.

[17]. Kim K. H, Park J. K, Hwang K. J, Kim S. H, Implementation of hybrid short-term load forecasting system using artificial neural networks and fuzzy expert systems, IEEE Trans. Power Syst., vol. 10, no. 3, pp. 1534–1539, Aug. 1995.

[18]. Senjyu T, Mandal P, Uezato K, and Funabashi T, Next day load curve forecasting using hybrid correction method. IEEE Trans. Power Syst., vol.20, No.1, pp.102-109, FEBRUARY 2005.

[19]. Blum L, Langley P, Selection of relevant features and examples in machine learning, Artif. Intell. 245-271, 1997.

[20]. Liu H, Motoda H, Feature selection for knowledge discovery and data mining, Kluwer Academic Publisher, Dordrecht, 1998.

[21]. Guyon I, Elisseeff A, An introduction to variable and feature selection, J. Mach. Learn. Res. 1157-1182, 2003.

[22]. Amjady N, Keynia F, Day-Ahead Price Forecasting of Electricity Markets by Mutual Information Technique and Cascaded Neuro-Evolutionary Algorithm, IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 24, NO. 1, FEBRUARY 2009.

[23]. Bernard. S, L. Heutte, S. Adam, Influence of hyperparameters on random forest accuracy, J.A. Benediktsson, J. Kittler, and F. Roli (Eds.): MCS 2009, LNCS 5519, pp. 171–180, 2009.

[24]. Liaw. A, M. Wiener, Classification and regression by randomForest, R News ISSN 1609-3631, Vol. 2/3, pp 18-22, December 2002.

[25]. Breiman L, Random Forest, Statistics Department University of California, Berkeley, CA 94720.

[26]. Amjady N, Keynia F, Electricity market price spike analysis by a hybrid data model and feature selection technique, Electr. Power Syst. Res. (2009), doi:10.1016/j.epsr.2009.09.015.

[27]. Peng H, Long F, Ding C, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) 1226–1238, 2005.

[28]. The New York Independent System Operator Web Site. [Online]. Available: http://www.nyiso.com.

[29]. PJM Web Site. [Online]. Available: http://www.pjm.com