Witold BARTKIEWICZ

Department of Computer Science, University of Łódź

Prediction Intervals for Short-Term Load Forecasting Neuro-Fuzzy Models

Abstract. In the paper the problem of estimation of the prediction intervals (error bars) for the family neuro-fuzzy Short-Term Load Forecasting (STLF) models is discussed. We investigate two neuro-fuzzy networks: Fuzzy Basis Function (FBF) Networks, and linear neuro-fuzzy model with Tagagi-Sugeno reasoning. The paper contains comparison of selected most important methods for error bars calculation (analytical delta method, and bootstrap), and discusses the obtained results in context STLF.

Streszczenie. W artykule zaprezentowane zostały metody wyznaczania przedziałów prognozy dla rodziny neuronowo rozmytych modeli krótkoterminowego prognozowania obciążenia sieci. Przebadane zostały dwa rodzaje sieci neuronowo-rozmytych: sieci Fuzzy Basis Function (FBF) i liniowe neuronowe modele rozmyte z wnioskowaniem typu Takagi-Sugeno. Artykuł obejmuje porównanie najistotniejszych metod szacowania przedziałów prognozy: analitycznej metody delta i bootstrapu), dyskutując wyniki w kontekście krótkoterminowych prognoz obciążenia sieci. (Metody wyznaczania przedziałów prognozy dla rodziny neuronowo rozmytych modeli krótkoterminowego prognozowania obciążenia sieci)

Introduction

Accurate prediction of the loads has a significant impact on economic and reliable operation of an electric power system. A lot of decisions and operating procedures in power companies require estimates of the energy demand in the future. In short time horizon, future loads are influenced mainly by their past values and weather factors (for instance temperature, pressure, humidity, etc.). The later relationship is commonly complex, implicit and nonlinear. Because of this short-term load forecasting (STLF) tools, based on nonlinear modelling methods, have ability to outperform classical short-term load forecasting models, especially during rapid changes in weather conditions. Good results have been achieved using neural networks [6-7], and neuro-fuzzy models [1][5][8].

In the paper the application of the family neuro-fuzzy models to short-term load forecasting problems is discussed. We investigate the problem of estimation of prediction intervals (error bars) for the energy demand forecasts. There were developed several method of error bars assignment for nonlinear regression models. They showed reasonable good accuracy in STLF tasks for neural networks models (multilayered perceptrons) [2-4], [6-7]. The goal of this paper is to verify their usefulness in case of neuro-fuzzy networks.

The other approaches to prediction intervals estimation for neuro-fuzzy systems follow general theory of uncertainty assessment in nonlinear models. They are based mainly in empirical error estimation on test data set [10] or analytical assessment on training set [9] (co called delta method). In this paper we investigate both kinds of approaches but in more complex form, with more exact results. The empirical approach considered here, in opposition to [10] is based on bootstrap approach, with resampling procedure of the test set. The delta method in [9] uses approximation of the model weights covariance matrix with outer product of first order derivatives of the model on training set patterns (so called outer product approximation or Levenberg-Marquard approximation). We have developed for this purposes full error Hessian algorithms for neuro-fuzzy networks. In case of STLF it gives significant improvement of the results.

The first considered model, Fuzzy Basis Function (FBF) Network is based on additive fuzzy logic system, with Gaussian fuzzy sets in premises of the rules, Larsen product aggregation rule, and utilizes crisp numbers in consequents of the rules. As a result, the FBF models are functionally equivalent Radial Basis Function neural networks, and show similar approximation capabilities. The equation of FBF network can be written as:

$$(1) \qquad y(x_1,...,x_n) = \frac{\sum_{i=1}^{K} b_i \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\frac{(x_j - a_{ij})^2}{\sigma_{ij}^2}\right)}{\sum_{i=1}^{K} \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\frac{(x_j - a_{ij})^2}{\sigma_{ij}^2}\right)}$$

where K is a number of rules, $x_1, …, x_n$ are input variables and y denotes an output variable. Constants a_{ij} and σ_{ij}, $j = 1, …, n$, $i = 1, …, K$, are Gaussian curve parameters, centers and widths, respectively, of the fuzzy sets A_{ij} in premises of the rules, and b_i, $i = 1, …, K$, are centroids of the fuzzy sets in rules consequences. Lets denote that FBF equation (1) uses so called Simplified Fuzzy Reasoning Rule, which means that the network is equivalent of the fuzzy system with constant crisp numbers (centroids) in rules consequences:

IF x_1 is A_{i1} AND... AND x_n is A_{in} THEN $y = b_i$
$$i = 1, …, K$$

The FBF model (1) can be considered as a kind of feedforward neural network. Computation of the partial derivatives, according parameters a_{ij}, σ_{ij}, b_i is relatively straightforward, and the model can be trained using least squares, gradient descent method, similar to back propagation learning algorithm for multilayered perceptron (MLP) networks, which was used in our case. Other learning methods, involving mixed clustering and least squares schemes, similar to the described below for Takagi-Sugeno networks, are also possible.

The FBF models are universal approximators, and it can be shown that from statistical point of view FBF model can be considered as Parzen-type mean-squared estimator of the conditional distribution probability density. Thus, fuzzy systems, given by equation (1) are capable to model complex nonlinear processes, like loads of the power network.

The next one, neuro-fuzzy model with linear Tagagi-Sugeno reasoning, extends the FBF concept, using rules consequents in form linear functions:

IF x_1 is A_{i1} AND... AND x_n is A_{in} THEN
$$y = m_{i0} + m_{i1}x_1 + … + m_{in}x_n , i = 1, …, K$$

where m_{ij}, $i = 1, \ldots, K$, $j = 0, \ldots, n$, are coefficients of the linear functions in consequents, other symbols like in FBF case. Also like in case of FBF models, assuming Gaussian type membership functions of fuzzy sets in premises of the rules, Larsen product aggregation operator and additive output defuzzification scheme, we can write the equation of the Takagi-Sugeno system, as follows:

$$v_i = m_{i0} + m_{i1}x_1 + \ldots + m_{in}x_n$$

(2)

$$y(x_1,\ldots,x_n) = \frac{\sum\limits_{i=1}^{K} v_i \exp\left(-\frac{1}{2}\sum\limits_{j=1}^{n}\frac{(x_j - a_{ij})^2}{\sigma_{ij}^2}\right)}{\sum\limits_{i=1}^{K} \exp\left(-\frac{1}{2}\sum\limits_{j=1}^{n}\frac{(x_j - a_{ij})^2}{\sigma_{ij}^2}\right)}$$

where constants a_{ij} and σ_{ij}, $j = 1, \ldots, n$, $i = 1, \ldots, K$, are also like for FBF network, Gaussian curve parameters, centers and widths, respectively, of the fuzzy sets A_{ij} in premises of the rules.

To obtain the forecasting neuro-fuzzy Takagi-Sugeno model (6) it is necessary to find the parameters a_{ij}, σ_{ij}, m_{ij}. It is possible to do the training process using gradient methods and error backpropagation algorithm, like for FBF network, but in this case we choose other learning scheme, performed in two stages:

• The first stage: calculation of the Gaussian membership functions parameters for fuzzy sets in premises of the rules. The centers a_{ij} of the Gaussian functions are determined through application of the clustering algorithm in the input space of the system. Calculated centroids of the clusters in training data, serve as centroids vectors of the multidimensional fuzzy sets $\mathbf{a}_i^*$ for each system rule. This way we may achieve the appropriate positioning of the all Gaussian curves in the input space. The width parameters σ_{ij} can be calculated according to relative distance between obtained clusters. But good results were obtained also with much straightforward approach, rely on simple partitioning of the each input variable domain

• The second stage: calculation of the linear functions coefficients in consequents of each rule. This task can be expressed after input transformations as generalized (curvelinear) least squares problem. The linear function coefficients are estimated, by the solving of the linear model for training data.

Uncertainty of the neuro-fuzzy STLF models

Lets denote by the $f(\mathbf{x}, \mathbf{w})$ output of the trained neuro-fuzzy network for a given input pattern $\mathbf{x}$, and parameters set $\mathbf{w}$. The parameters of the neuro-fuzzy networks, it means centres and width of the Gaussian functions in premises, and the coefficient in consequents of the rules, usually are trained using least squares method, or two-stage mixed learning schemes, involving input-space clustering and curve-linear regression. As a result obtained model approximates the conditional expected value of the target predictive load distribution, conditioned on the input pattern.

Such regression-type point predictors do not evaluate all necessary aspects of the forecasting problem. The decision maker, using neuro-fuzzy STLF model, in order assessing the uncertainties involved in the prediction, needs additional information obtained from the entire conditional target distribution of the forecasted load process. We discuss the topic of the conditional variance assessment for the

prediction, around the new data point, in case of neuro-fuzzy predictor.

In the paper we apply the obtained results only to the problem of the prediction intervals (error bars), for the energy demand forecasts, but they can be used also to other decision risk analyses, like the optimal ordering strategies on the market development, calculation risk margins for the load reserves, making risk-free optimal decisions, etc.

There are several sources of uncertainty, we should take into account, when we try to estimate prediction intervals for neuro-fuzzy predictor.

As it was stated before, the neuro-fuzzy network approximates the expected value of the conditional target distribution. The first source of uncertainty of the prediction arises from the error of this approximation. Usually it is called model bias. The second source uncertainty, model variance, results mainly from sampling variation. The training set $D = \{\mathbf{x}_k, y_k\} = \{(x_{k1}, \ldots, x_{kn}), y_k\}$, $k = 1, \ldots, N$ is sampled from the underlying general population. There is certain variability associated with this process. Different data sets, sampled from the same underlying relationship, result in different sets of parameters (weights), and in consequence in different predictions.

Development of the proper neuro-fuzzy forecasting model requires a trade-off between model bias and variance. Bias is due to a regression function having insufficient flexibility to model the data adequately enough. However reducing the bias by increasing the complexity of the model or training efforts results in increasing the variance (well known problem of overtraining). A model better fitted to training data is more sensitive to sampling variability. Neuro-fuzzy networks with their good approximation capabilities are well known as low bias and high variance estimators. There are other reasons of bias, like for instance too late or to early stopping of the training algorithm. But for proper developed neuro-fuzzy model the bias should be relatively small comparing the variance. Because of this further in this paper we will assume, that the model is unbiased.

The intervals resulted from the training set sampling variation, obtained for the parameters and model output, are usually referred as confidence intervals. To obtain the prediction intervals for the new input pattern $\mathbf{x}$, we should consider also other important source of uncertainty of the forecasts. It is connected with intrinsic and irreducible random error. The relationship between dependent and explanatory variables, contains the stochastic component, so called target noise. Because of the independence of the network parameters and the error term, variance $\sigma_y^2(\mathbf{x})$ of the target distribution can be decomposed into two following components:

(3) $$\sigma_y^2(\mathbf{x}) = \sigma_w^2(\mathbf{x}) + \sigma_\varepsilon^2(\mathbf{x})$$

where $\sigma_w^2(\mathbf{x})$ denotes the variance of the model, associated with weights uncertainty, and $\sigma_\varepsilon^2(\mathbf{x})$ means the variance of the random error.

It can be shown, that for STLF neuro-fuzzy models, there are some empirical backgrounds, to assume Gaussian model of the error probability distribution. In considered in this paper forecasting problems, testing of the model residuals allowed to confirm this assumption (see discussion of the target noise variance estimation, in the next chapter). So, further we will assume that conditional probability distribution of the forecasted load y, for a given input pattern $\mathbf{x}$, will have character of the normal distribution

$N(f(\mathbf{x}, \mathbf{w}), \sigma_y(\mathbf{x}))$, with expected value given by the model output (load forecast) $f(\mathbf{x}, \mathbf{w})$, and standard deviation $\sigma_y(\mathbf{x})$ defined by (3), with probability density function:

$$(4) \qquad p(y / \mathbf{x}) = \frac{1}{\sigma_y(\mathbf{x})\sqrt{2\pi}} \exp\left(- \frac{(y - f(\mathbf{x}, \mathbf{w}))^2}{2\sigma_y^2(\mathbf{x})}\right)$$

Further consideration in this paper are based upon a normal (Gaussian) distribution of the load forecast. As it was noted, in our research, there were basis for such assumption, but it should be verified in any individual case.

Usually for a distribution of the forecasted load (4), α-prediction interval given is by:

$$(5) \qquad f(\mathbf{x}, \mathbf{w}) \pm t_{N-p}^{1-\alpha/2} \sigma_y(\mathbf{x})$$

where $t_{N-p}^{1-\alpha/2}$ denotes percentile of the t-Student distribution.

The number degree of freedom $N - p$ is equal difference between number training patterns N and number of parameters p. But in our case, because of relatively large sizes of training sets (more than 500 training samples in each of performed experiments), there is not necessity to use t-Student distribution, and prediction interval may by calculated directly from normal distribution (4):

$$(6) \qquad f(\mathbf{x}, \mathbf{w}) \pm Q_{N(0,1)}((1+\alpha)/2)\sigma_y(\mathbf{x})$$

where $Q_{N(0,1)}((1+\alpha)/2)$ denotes $(1+\alpha)/2$ quantile of standard normal distribution.

Assessment of the forecast variance

Lets remind that according to (3), the forecast variance $\sigma_y^2(\mathbf{x})$, consists of two basic components: $\sigma_w^2(\mathbf{x})$ - model resulting from parameters uncertainty, and noise variance $\sigma_\varepsilon^2(\mathbf{x})$. In this chapter we will discuss assessment methods of those components.

A. Noise variance estimation

Reasoning about output uncertainty of the regression model in ordinary least squares method, usually uses assumption of constant value of the target noise variance, $\sigma_\varepsilon^2(\mathbf{x}) = \sigma_\varepsilon^2$. The estimator of σ_ε^2 is based on standard residual error of the model:

$$(7) \qquad \sigma_\varepsilon^2 \approx \frac{1}{N-p} SSE$$

where SSE is sum of squares of the neuro-fuzzy network residuals.

However our experiences show, that in case non-linear neural and neuro-fuzzy STLF models, assumption of constant σ_ε^2 variance of the target noise, usually doesn't hold. We obtained much better results, using additional input dependent noise model $\sigma_\varepsilon(\mathbf{x})$, conditioned on the forecast input pattern $\mathbf{x}$, and obtained with additional neural network (multilayered perceptron).

Actually using input dependent variance (standard deviation) model of random noise, was necessary for confirmation of the normality of predictive load distribution.

If we denote the training set for the neuro-fuzzy STLF model by $D = \{\mathbf{x}_k, y_k\} = \{(x_{k1}, \ldots, x_{kn}), y_k\}$, $k = 1, \ldots, N$, the additional neural network supplying standard deviation of the random noise $\sigma_\varepsilon(\mathbf{x})$, was trained using following set of patterns:

$$(8) \qquad D = \{x_k, | e_k |\}, k = 1, \ldots, N$$

where as target values for the training of the additional networks $|e_k| = |y_k - f(\mathbf{x}_k, \mathbf{w})|$, $k = 1, \ldots, N$, are used the absolute values of the basic forecasting model training residuals.

B. Assessment of the model variance from parameters, with bootstrap

We discuss two main approaches to calculation of the output variance, resulted from uncertainty of the parameters $\sigma_w^2(\mathbf{x})$, developed for non-linear regression models:
- The first one is based on bootstrap.
- The second, delta method, is based on the estimation of the covariance matrix of the network weights.

Both methods were created mainly for neural network predictors, and showed good results in STLF tasks, in case multilayered perceptron (MLP) networks [2,3][6-7]. In this paper we will analyze their application to the FBF, and Takagi-Sugeno neuro-fuzzy models.

The bootstrap is pure empirical approach, based on resampling of the learning set, drawing (with replacement) from learning data patters certain number of samples, and training for each sample the separate neuro-fuzzy network. The model variance for a given input pattern, resulting from weights uncertainty, is estimated using variance of the model output over the network representations for different samples.

More detailed speaking, we have utilized so called pairs sampling approach. From training data set $D = \{\mathbf{x}_k, y_k\} = \{(x_{k1}, \ldots, x_{kn}), y_k\}$, $k = 1, \ldots, N$, we draw with replacement (with uniform probability distribution) B bootstrapped samples D_i, each of the consisted of N training patterns. For every samples D_i, we train the appropriate neuro-fuzzy model $f(\mathbf{x}, \mathbf{w}_i)$, $i = 1, \ldots, K$. Output variance from parameters $\sigma_w^2(\mathbf{x})$, may be calculated by:

$$(9) \qquad \sigma_w^2(\mathbf{x}) = \frac{1}{B-1}\sum_{i=1}^{B}(f(\mathbf{x}, \mathbf{w}_i) - f_{avg}(\mathbf{x}, \cdot))^2$$

where $f_{avg}(\mathbf{x}, \cdot)$ denotes average of the forecasts, obtained from all models $f(\mathbf{x}, \mathbf{w}_i)$, $i = 1, \ldots, K$:

$$(10) \qquad f_{avg}(\mathbf{x}, \cdot) = \frac{1}{B}\sum_{i=1}^{B} f(\mathbf{x}, \mathbf{w}_i)$$

Finally in this case, the estimator for variance of the load forecast distribution (4), conditioned on input pattern $\sigma_y^2(\mathbf{x})$, given is by equation (3), where $\sigma_\varepsilon^2(\mathbf{x})$ is random noise variance discussed in previous point, and $\sigma_w^2(\mathbf{x})$ is obtained from (9).

C. Assessment of the model variance from parameters, with delta method

The delta method it is analytical approach, based on local linearization of the neuro-fuzzy network in the neighborhood of optimal (trained) parameters vector $\mathbf{w}^*$, using Taylor expansion:

$$(11) \qquad f(\mathbf{x}, \mathbf{w}) = f(\mathbf{x}, \mathbf{w}^*) + \mathbf{g}(\mathbf{x}, \mathbf{w}^*)\Delta\mathbf{w}$$

where $\mathbf{g}(\mathbf{x}, \mathbf{w}^*)$ is the gradient vector of the network with respect to the weights (parameters), and $\Delta\mathbf{w} = \mathbf{w} - \mathbf{w}^*$. Hence, using error propagation law, we can write formula for the output variance from for a given input pattern $\mathbf{x}$:

$$(12) \qquad \sigma_w^2(\mathbf{x}) = \mathbf{g}(\mathbf{x},\mathbf{w})^{\mathrm{T}} \mathbf{C_w} \mathbf{g}(\mathbf{x},\mathbf{w})$$

where, $\mathbf{g}(\mathbf{x},\mathbf{w}^*)$ denotes like in (11) the gradient of the network, and $\mathbf{C_w}$ is covariance matrix of the network's weights.

The covariance matrix $\mathbf{C_w}$ can be approximated using inverse of the Hessian matrix $\mathbf{H}$ of the model error, with respect to the weights:

$$(13) \qquad \mathbf{C_w} = \sigma_\varepsilon^2(\mathbf{x})\mathbf{H}^{-1}$$

where $\sigma_\varepsilon^2(\mathbf{x})$ is random noise variance obtained by the additional neural network, as discussed in point A.

Substituting (13) in (3) we will obtain final delta method formula for the model output variance for a given input pattern $\mathbf{x}$:

$$(13) \qquad \sigma_y^2(\mathbf{x}) = \sigma_\varepsilon^2(\mathbf{x}) + \sigma_\varepsilon^2(\mathbf{x})\mathbf{g}(\mathbf{x},\mathbf{w}^*)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{g}(\mathbf{x},\mathbf{w}^*)$$

Obtained results and conclusions

In the first testing task, we have applied both approaches, to FBF and Takagi-Sugeno (TS) neuro-fuzzy networks, using the problem of two-day ahead, hourly energy demand forecasting, on the peak hour. The width of the obtained with both methods prediction intervals was tested on independent data set, and empirical frequencies of the actual loads inside the estimated interval, verified against assumed theoretical expected probability levels α, used in calculations.

Results presented in Table 1, show approximately correct frequencies, similar for either both types of networks, and both methods. The width of intervals calculated for several typical probability levels α, is appropriate enough, to gave analogous frequencies levels. So the empirical verification of the prediction intervals, estimated using bootstrap and delta method show good, results in both cases.

Table 1. Frequencies of obtained prediction intervals for energy demand forecasts on the peak hour

Prob. level	Delta method		Bootstrap	
	TS	FBF	TS	FBF
80%	76.95%	81.23%	77.11%	81.06%
85%	83.53%	85.83%	83.69%	85.68%
90%	88.25%	89.80%	88.11%	89.64%
95%	92.49%	93.64%	92.49%	93.94%

Table 2. Frequencies of obtained prediction intervals testing for daily energy demand forecasting

Prob. level	Delta method		Bootstrap	
	MLP	FBF	MLP	FBF
80%	81.24%	81.77%	81.84%	80.06%
85%	85.63%	86.16%	85.80%	84.36%
90%	89.57%	90.25%	89.86%	89.34%
95%	93.99%	93.25%	93.99%	93.49%

The second testing task, was one-day ahead daily energy demand forecasting problem. In this case we compare prediction intervals estimated with delta method and boostrap, for FBF and MLP (multilayered peceptron)

neural network [4]. As we can see in Table 2, also in this case results obtained with both approaches for FBF network are approximately correct. Either delta method and bootstrap, show similar accuracy for neuro-fuzzy model, like for neural network.

In conclusion, it can be stated, that for considered in our paper short-term load forecasting problems, performed researches, showed correct accuracy of the prediction intervals calculated both with delta method and bootstrap-based approach. Both methods cope with the problem from completely different directions. Bootstrap estimate is pure empirical in the nature, based only on data and behaviour of the model itself. From this point of view, it may be considered as a much more reliable, than based on the approximate theory for linear regression models, with many only roughly fulfilled assumptions, delta method.

But empirical verifications did not show this difference. Both methods worked very similar. It should be noted, that delta method requires much less computational efforts than bootstrap estimator. And it is the reason, that for practical STLF tasks, we recommend rather this approach as a first-choice method.

REFERENCES
[1] Bartkiewicz W., Neuro-fuzzy approach to short-term electrical load forecasting. *Proc. of the International Joint Conference on Neural Networks, IJCNN2000*, Como, Italy, July 24-27, 2000, vol. VI, 229-234.
[2] Bartkiewicz W., Confidence intervals prediction for the short-term electrical load neural forecasting models, *Elektrotechnik und Informationstechnik*, No. 1(117), 2000, 8-12.
[3] Bartkiewicz W., Error Bars for Short Term Load Forecasting Neural Networks Models. *Present-Day Problems of Power Engineering (APE '2001)*, Gdańsk - Jurata, 2001, vol. III, 75-82.
[4] Bartkiewicz W., Metody określania niepewności prognoz krótkoterminowego obciążenia sieci dla modeli neuronowych i neuronowo-rozmytych. *Rynek Energii*, nr 1 (92), luty 2011, s. 41 – 46.
[5] Bartkiewicz W., Short-Term Load Forecasting with Neuro-Fuzzy Models. *Present-Day Problems of Power Engineering (APE '2011)*, Gdańsk - Jurata, 2011, (to appear).
[6] Bartkiewicz W., Gontar Z., Matusiak B., Zieliński J.S., Short-term load forecasting in market environment, *Proc. of III-d IEE Mediterranean Conference and Exhibition on Power generation, Transmission, Distribution and Energy Conversion (Med Power)*, Athens. Nov 4 – 6, 2002.
[7] Bartkiewicz W., Gontar Z., Zieliński J.S., Bardzki W.: Uncertainty of the short-term load forecasting in utilities, *Proc. of the International Joint Conference on Neural Networks, IJCNN2000*, Como, July 24-27, 2000, vol. VI, 235-240,
[8] Butkevych O.F., Pawłowskiy W.W., Bartkiewicz W., Zieliński J.S., Hybrid systems in power systems solving. *Technical Electrodynamics*, Institute of Electrodynamics National Academy of Sciences, Kiev 2002, 77-82.
[9] Khosravi, A., Nahavandi, S. and Creighton, D., Prediction Interval Construction and Optimization for Adaptive Neuro Fuzzy Inference, *IEEE Transactions on Fuzzy Systems*, (to appear).
[10] Mencar C., Castellano G., Fanelli M., Deriving Prediction Intervals For Neuro-Fuzzy Networks. *Mathematical and Computer Modelling*, vol. 42 (2005), issue 7-8, 719-726.

Author: dr Witold Bartkiewicz, Uniwersytet Łódzki, Katedra Informatyki, ul. Matejki 22/26, 90-237 Łódź, E-mail: wbartkiewicz@wzmail.uni.lodz.pl.