Tomasz CIECHULSKI[1], Stanisław OSOWSKI[1,2]

Military University of Technology, Faculty of Electronics (1), Warsaw University of Technology, Faculty of Electrical Engineering (2)

doi:10.15199/48.2017.10.13

Local dynamic integration of ensemble of predictors in load forecasting

Abstract. *The paper shows the new approach to integration of an ensemble of neural predictors in load forecasting. In opposite to classic integration method built upon weighted averaging of every single predictor results this integration method uses only the results of one predictor which was the best on the input data of the learning vectors from the past, which were closest to the actual excitation. Thanks to this the result of ensemble is never worse than the best unit in ensemble. The results of 24-hour ahead prediction of the daily load in small power system have confirmed the efficiency of the proposed solution.*

Streszczenie. *Artykuł przedstawia nowe podejście do integracji zespołu predyktorów neuronowych w zadaniu prognozowania godzinnych obciążeń dobowych z wyprzedzeniem 24-godzinnym. W metodyce tej do predykcji używany jest tylko jeden – najlepszy predyktor dla analizowanej doby. Konkretny wektor obciążeń z danych uczących wraz z najbardziej dokładną odpowiadającą mu siecią neuronową wyłonioną w trybie uczenia wybierany jest na podstawie najmniejszej odległości euklidesowej badanego wektora w trybie testującym. Wyniki badań numerycznych potwierdzają wyższość prezentowanej metody nad rozwiązaniami klasycznymi predykcji. (**Integracja dynamiczna zespołu predyktorów w zastosowaniu do prognozowania obciążeń elektroenergetycznych**).*

Keywords: load forecasting, neural networks, ensemble of predictors, local dynamic integration.
Słowa kluczowe: prognozowanie obciążeń, sieci neuronowe, zespół predyktorów, integracja dynamiczna.

Introduction

The work shows a new approach to integration of an ensemble of neural predictors in load forecasting in a small power system. The group of predictors is composed of many individual neural networks, trained on the same set of learning data. Each network is of different type and generates its own response, which might vary from unit to unit.

Conventional integration [5] is based on weighted averaging of every predicted results. The units generating more accurate results are assigned higher priority represented by the proper weighs in averaging process. However, in such system at different quality of predicting units in ensemble, the result of group can be worse than result of the best predictor (worse units have bad influence on the ensemble result).

Local dynamic integration principle

The new concept of integration has been presented in this paper. It is so called local dynamic integration. At the beginning every predictor is subject to classical learning process [1, 2, 3, 6]. Then the learning error committed by each predictor is computed for all observations from the learning set.

In the testing mode (the real prediction on the data not taking part in learning), the actual input vector $\mathbf{x}_t$ is compared to all vectors $\mathbf{x}_u$ used in learning. The nearest to it learning vectors $\mathbf{x}_u$ are selected, according to their Euclidean distance. In the extreme case the selected set can be limited to only one closest vector. Then, the prediction errors committed by all members of ensemble are computed for the chosen set of vectors $\mathbf{x}_u$. The network generating the lowest error on the closest learning data will be used in prediction using the actual testing vector $\mathbf{x}_t$ as an excitation. Other predictors do not take part in generating the results. It means the networks, which were worse for the data closest to the actual excitation, do not influence the final result of an ensemble. The integration is dynamic and local, since for each testing data different, locally the best predictor from the whole ensemble, is chosen to generate the forecast. The other testing vectors $\mathbf{x}_t$ can use quite different predicting units.

Three different neural networks will be used in an ensemble. They include multilayer perceptron (MLP), radial basis function network (RBF) and support vector machine (SVM) [3, 4, 5]. The process of 24-hour ahead load prediction for the next day by using the local dynamic integration of predictors will be performed according to the following algorithm:

1. Learning MLP, RBF and SVM neural networks on the same learning data set. As a result of learning we get the predicted vectors $\mathbf{y}_i(\mathbf{x}_i)$ on the learning data for all members of ensemble. The prediction error is defined in the form of an Euclidean distance between $\mathbf{y}_i(\mathbf{x}_i)$ and the appropriate destination $\mathbf{d}(\mathbf{x}_i)$

$$\varepsilon_i = ||\mathbf{y}(\mathbf{x}_i) - \mathbf{d}(\mathbf{x}_i)|| \tag{1}$$

and is calculated for each individual predictor.

2. The prediction of the load for the new testing vector $\mathbf{x}_t$ not taking part in learning is carried out in the following way:

a) compute the differences of $\mathbf{x}_t$ to every vector from the learning set. The differences are defined using the Euclidean norm $||\mathbf{x}_t - \mathbf{x}_{ui}||$;

b) find the learning vectors which are the nearest to the testing vector $\mathbf{x}_t$ according to this distance;

c) calculate the prediction errors committed by the individual predictors for the chosen set of learning data $\mathbf{x}_u$;

d) chose the neural network of the smallest error in the learning data to predict the load for the input testing vector $\mathbf{x}_t$. This result will be assumed as the final result of an ensemble. In this way only one member of an ensemble is responsible for final prognosis

In general solution different population of learning vectors might be also considered. The most popular is the use of single vector $\mathbf{x}_u$ for which the prediction is done. However, in the case when different number of closest learning vectors is chosen, take the best neural predictors for each of them and use these networks in the actual prediction for the testing vector $\mathbf{x}_t$. The resulting forecast will be the average of predictions made by all used networks.

Application example

Three neural networks: MLP, RBF and SVM have been learned using the data base of power consumption in a small power system in Poland The data of the year 2004 have been used in this process.

Let us predict the 24-hour load vector for exemplary day of 2nd December 2005. The normalized input vector for this prediction, limited to the power load of the previous day, is as follows:

$\mathbf{x}_t$ = [0,505 0,462 0,444 0,442 0,453 0,499
0,614 0,695 0,737 0,764 0,776 0,779
0,778 0,782 0,768 0,759 0,814 0,837
0,828 0,805 0,766 0,721 0,666 0,573]

According to the dynamic approach to integration we look first to the nearest vector $\mathbf{x}_u$ from the learning set. The Euclidean distance was smallest to the input vector used in load prediction for 24th November 2004. Its specific form is as follows

$\mathbf{x}_u$ = [0,506 0,463 0,445 0,443 0,457 0,504
0,614 0,691 0,733 0,761 0,773 0,775
0,775 0,780 0,766 0,757 0,814 0,841
0,838 0,819 0,778 0,726 0,668 0,575]

The Euclidean difference between these vectors is equal:

$\varepsilon = ||\mathbf{x}_t - \mathbf{x}_u|| = 0{,}0249$.

For this learning vector the following MAPE errors have been committed by the ensemble units

- $MAPE_{SVM}$ = 0,59%,
- $MAPE_{RBF}$ = 0,93%,
- $MAPE_{MLP}$ = 1,01%.

It is clearly visible that the most accurate forecast was generated by the SVM network and so this network was used in 24-hour load prediction for the tested day. The obtained mean error calculated for 24 hours of the day was equal 1,36%. The graphical results, showing the 24-hour prediction curve (-·-·) and the real (true) load pattern in examined day (–) are depicted in Fig. 1.

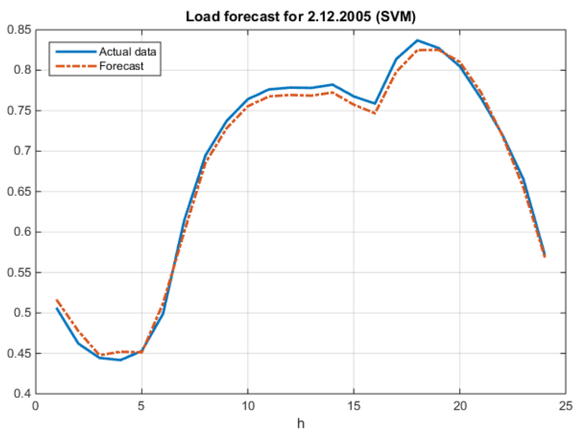

Fig. 1. The 24-hour load pattern for chosen day generated by SVM network

The use of other types of networks resulted in a worse quality of prediction. In these cases the mean value of errors for 24-hours were as follows: 1,45% for RBF and 1,54% for MLP networks.

In the dynamic integration of an ensemble we can also use many learning vectors, which are nearest to the testing vector. For each case the best predictor is chosen and then used for prediction. Final forecast is the average of the results of the applied predictors.

The use of few different predictors in dynamic integration is in some cases beneficiary because of nonlinear relationship between the prediction errors for the particular hours and the Euclidean difference $||\mathbf{x}_t - \mathbf{x}_u||$. The averaging process leads usually to the relaxation of these negative factors.

Statistical results of dynamic integration

To determine and compare the statistical results of application of the dynamic integration of an ensemble the prediction experiments for two years 2004 and 2005 have been carried out. The data of 2004 have been used only for learning these three neural networks and 2005 left for testing purposes.

Table 1 presents the statistical results in the form of learning MAPE errors committed by SVM, MLP and RBF in one month of the 2004 year. The data refer to the results of load prediction for all days of January.

Table 1. MAPE learning errors committed by three neural networks

January 2004							
Day	MAPE error [%]			Day	MAPE error [%]		
	MLP	RBF	SVM		MLP	RBF	SVM
1.	1,03	3,08	1,76	17.	0,97	0,44	1,36
2.	2,77	3,71	1,67	18.	1,38	2,06	1,24
3.	2,35	1,51	1,18	19.	2,41	0,89	1,31
4.	2,02	2,85	2,34	20.	3,78	1,97	1,26
5.	1,08	1,77	1,45	21.	2,38	2,71	1,63
6.	1,13	0,87	1,84	22.	2,81	0,77	1,91
7.	1,60	0,42	2,96	23.	2,29	2,26	1,43
8.	2,59	3,27	1,38	24.	1,60	0,60	1,39
9.	1,67	3,16	1,53	25.	1,12	2,33	1,83
10.	2,29	1,03	1,36	26.	1,45	5,61	1,96
11.	1,97	1,46	1,31	27.	3,91	2,40	2,38
12.	3,72	2,37	2,11	28.	2,19	3,23	2,09
13.	1,26	3,16	1,63	29.	2,45	3,06	2,18
14.	0,75	4,27	1,86	30.	2,25	0,66	1,67
15.	2,32	3,72	0,98	31.	2,57	2,83	1,81
16.	2,71	4,17	2,41	**Mean**	**2,09**	**2,34**	**1,72**
Mean error of the best network for each day: **1,32%**							

The profit of using the best predictors is evident. The mean MAPE errors committed by the networks were as following: MLP – 2,09%, RBF – 2,34% and SVM – 1,72%. However, using only the best predictors for each day leads to reduction of this error to only 1,32%.

Similar improvements have been observed for other months of the year. Table 2 presents the average monthly values of MAPE learning errors using only the best neural networks used in prediction of the daily load in the year 2004.

Table 2. Statistical results of MAPE learning errors of the best neural network in each day for all months of the year 2004

Month	MAPE learning error	Month	MAPE learning error
January	1,32%	July	1,41%
February	1,42%	August	1,42%
March	1,64%	September	1,56%
April	1,48%	October	1,78%
May	1,79%	November	1,80%
June	1,44%	December	1,69%

The percentage improvement of prediction accuracy thanks to application of the presented methodology measured for learning data for the whole year is more than 30%.

The real assessment of the presented method may be done on the testing data, not taking part in learning. The learned neural networks have been used in prediction process for the load in the year 2005. The dynamic integration used only one, the nearest vector to the tested vector, so only one predictor was employed in producing final forecast of an ensemble. The exemplary results of integration for one chosen month of the year 2005 are presented in table 3.

Table 3. MAPE testing errors committed by an ensemble for all days of January 2005

January 2005					
Day	MAPE error [%]	Best network	Day	MAPE error [%]	Best network
1.	2,63	SVM	17.	1,33	SVM
2.	1,15	SVM	18.	1,48	SVM
3.	1,16	SVM	19.	1,41	SVM
4.	1,04	SVM	20.	1,59	SVM
5.	1,93	MLP	21.	2,16	SVM
6.	1,36	SVM	22.	1,66	SVM
7.	2,05	RBF	23.	2,87	MLP
8.	1,06	SVM	24.	1,43	RBF
9.	1,22	SVM	25.	1,51	SVM
10.	1,41	SVM	26.	1,96	SVM
11.	1,03	SVM	27.	1,39	SVM
12.	1,04	SVM	28.	1,28	SVM
13.	1,45	SVM	29.	1,40	SVM
14.	1,49	SVM	30.	1,89	SVM
15.	1,37	SVM	31.	1,55	SVM
16.	1,34	SVM	**Mean**	**1,54**	

It is evident that SVM was chosen as the best in the majority of times. However, four times the other networks have been also selected as the better predictors.

Table 4 presents the statistical results of MAPE testing errors for every month of the year 2005. In most cases the best results of prediction were due to application of SVM. The participation of other networks (MLP and RBF) in creation of final forecast is assessed on the level of 15%.

Table 4. Statistical results of MAPE testing errors for the year 2005 using dynamic integration of predictors in an ensemble

Month	MAPE testing error	Month	MAPE testing error
January	1,54%	July	1,74%
February	1,62%	August	1,77%
March	1,78%	September	1,90%
April	1,72%	October	1,96%
May	1,89%	November	2,04%
June	1,73%	December	1,86%
The mean MAPE for the whole year is **1,80**%			

Table 5. Statistical results of MAPE testing errors for all months of the year 2005 using single SVM predictor

Month	MAPE testing error	Month	MAPE testing error
January	1,93%	July	2,24%
February	2,21%	August	2,17%
March	2,36%	September	2,29%
April	2,33%	October	2,46%
May	2,31%	November	2,36%
June	2,23%	December	2,09%
The mean MAPE for the whole year is 2,25%			

It is interesting to compare the results of an ensemble to the results of application of only the best SVM predictor. Table 5 depicts the appropriate statistical results in such case. The presented results confirm the superiority of an ensemble application.

Conclusion

The paper has presented the new approach to integration of the ensemble members in the task of load forecasting in small power system. In classic approach all units belonging to the ensemble take part in making the final decision. In hereby presented dynamic method, the final prediction of the group is made by the network which has had the best fit on the similar learning vector in the past. Only one network which generated the best result, measured by the MAPE error, is responsible for the final forecast. In this way we can avoid disadvantageous situation when the worst unit in the ensemble worsens the forecast accuracy in relation to the best individual result.

Statistically, according to the carried out numerical experiments the vast majority of chosen networks was the SVM which dominated over other predictors. However, its direct (single) application in forecasting process has led to worse statistical results.

Authors: *dr inż. Tomasz Ciechulski, Military University of Technology, Institute of Electronic Systems, Division of Measurement and Information Systems, ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa,*
E-mail: tomasz.ciechulski@wat.edu.pl;
prof. dr hab. inż. Stanisław Osowski, Warsaw University of Technology, Institute of Theory of Electrical Engineering, Measurement and Information Systems, ul. Koszykowa 75, 00-662 Warszawa and Military University of Technology, Institute of Electronic Systems, Division of Measurement and Information Systems, ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa
E-mail: sto@iem.pw.edu.pl.

REFERENCES

[1] Ciechulski T., Osowski S., *Badanie jakości predykcji obciążeń elektroenergetycznych za pomocą sieci neuronowych SVM, RBF i MLP*, Przegląd Elektrotechniczny, no. 8/2014, pp. 148-151
[2] Li S., Wang P., Goel L., *A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection*, IEEE Transactions on Power Systems, vol. 31, no. 3, pp. 1788-1798, 2016
[3] Lin C.J., Chang C.C., *Libsvm: a library for support vector machines*, https://www.csie.ntu.edu.tw/~cjlin/libsvm
[4] MathWorks, *Matlab manual user's guide*, Natick, 2016
[5] Osowski S., Siwek K., Szupiluk R., *Ensemble neural network approach for accurate load forecasting in the power system*, Applied Mathematics and Computer Science, 2009, vol. 19, No. 2, pp. 303-313
[6] Price P.N., *Methods for analyzing electric load shape and its variability*, Environmental Energy Technologies Division, 2010