Ryszard SZUPILUK[1], Tomasz ZĄBKOWSKI[2]

Warsaw School of Economics (1), Warsaw University of Life Sciences (2)

Noise identification for ICA ensemble predictors

***Abstract**. In this paper we present a novel method for integration the prediction results by finding common latent components via independent component analysis. The latent components can have constructive or destructive influence on particular prediction results. After the elimination of the deconstructive signals we rebuilt the improved predictions. We check the method validity on the electricity load prediction task.*

***Streszczenie.** W artykule przedstawiono nową metodę pozwalającą na łączenie wyników predykcji poprzez poszukiwanie ukrytych wspólnych składowych przy zastosowaniu procedury analizy składowych niezależnych. Składowe ukryte mogą mieć pozytywny lub negatywny wpływ na wyniki predykcji. Po wyeliminowaniu składowych niekorzystnych poprawione zostały wyniki predykcji. Poprawność metody sprawdzono na przykładzie predykcji zapotrzebowania na energię elektryczną. (**Identyfikacja szumów z wykorzystaniem metody ICA w kontekście agregacji**).*

Keywords: noise identification, ICA, electric load forecasting
Słowa kluczowe: identyfikacja szumów, ICA, prognozowanie obciążenia elektrycznego

Introduction

Forecasting the electricity demand is one of the most important areas of the research in energetics. Companies in the industry need both, short-term forecasts (minutes, hours or days) and long-term (up to several years). The importance of the former mentioned, increases with the development of competition and free market mechanisms in the electricity markets. Forecasting of the demand, it is essential for power sector, but in fact is very difficult. Firstly, due to the fact that the time series shows seasonal effects (daily, weekly and on annual basis). Secondly, due to external factors that has significant impact on the demand (i.e. meteorological factors) [4].

Moreover, there is growing influence of financial markets on energy markets. In recent years a significant increase in the inflow of financial investments in commodity derivatives markets (including electricity) has been observed. For example, in the years 2003-2008 in Europe, institutional investors increased their investments in the commodity market from 13 billion in 2003 to the amount of oscillating between 170 and 205 billion in 2008 [8]. As the financial crisis interrupted this trend, the financial situation of many markets in 2010, came close to a peak in 2008, and was even better. Although still under discussion on the relative influence of various factors on the prices of goods, it is clear that changes in prices of various commodities markets are closely linked, and that commodity markets are in close relation with the financial markets.

It is not difficult to create a short-term forecast with an error of a few percent. However, the financial costs of such an error are so high that the number of researches were undertaken in order to reduce even a fractional part of it. Therefore, in this article we develop an ICA approach for ensemble predictions. Its main idea is based on decomposition of the prediction results into underlying independent components. Some of these components may be associated with the prediction (the real value) and some of them can be treated as noise or interference. Elimination of noises, termed as destructive components, should result in prediction improvement. The ICA approach, represents chosen data decomposition from wide set of blind signal separation methods, but other techniques like SOS BSS, smooth component analysis, sparse component analysis can be applied either.

The term ensemble or aggregation is a consequence of the fact that the final result is a combination of individual results from different models. In opposite, to other popular ensemble methods like bagging or boosting, there will be no assumptions to the form of aggregated models nor to criteria for model assessment. In other words, we can aggregate models (more specifically, the results of their prediction) regardless to specific criterion [3, 5, 12].

From an operational point of view, we can say, that the goal of this method, is to remove from the prediction the noise or any disruption that has physical nature. This can be described as a filtration of prediction results, but taking into account the multiplicity of models, this approach is rather designed to provide information about the way in which we can effectively extract the noise.

Prediction results improvement

We assume, that after the learning process, each prediction result includes two types of latent components: constructive, associated with the target, and destructive, associated with the inaccurate learning data, individual properties of models, missing data, not precise parameter estimation, distribution assumptions etc. Let us assume there is m models. We collect the results of particular model in column vector x_i, $i=1,...,m$ and treat such vectors as multivariate variable $X=[x_1,x_2,...,x_m]^T$, $X \in R^{m\times N}$, where N means the number of observations. We describe the set of latent components as $\mathbf{S}=[\hat{\mathbf{s}}_1,\hat{\mathbf{s}}_2,...,\hat{\mathbf{s}}_k,\mathbf{s}_{k+1},\mathbf{s}_n]^T$, $S \in R^{m\times N}$, where $\hat{\mathbf{s}}_j$ denotes constructive component and $\mathbf{s}_i$ is destructive one [9]. For simplicity of further considerations we assume $m=n$. Next, we assume the relation between observed prediction results and latent components as linear transformation

(1) $$\mathbf{X}=\mathbf{AS}$$

where matrix $\mathbf{A} \in R^{n\times n}$ represents the mixing system. The (1) means matrix X decomposition by latent components matrix S and mixing matrix $\mathbf{A}$.

Our aim is to find the latent components and reject the destructive ones (replace them with zero). Next we mix the constructive components back to obtain improved prediction results as

(2) $$\hat{\mathbf{X}}=\mathbf{A}\hat{\mathbf{S}}=\mathbf{A}[\hat{\mathbf{s}}_1,\hat{\mathbf{s}}_2,...,\hat{\mathbf{s}}_k,\mathbf{0}_{k+1},...,\mathbf{0}_n]^T$$

The replacement of destructive signal by zero is equivalent to putting zero in the corresponding column of A. If we express the mixing matrix as $\mathbf{A}=[a_1,a_2,...,a_n]$ the purified results can be described as

(3) $$\hat{\mathbf{X}}=\hat{\mathbf{A}}\mathbf{S}=[\mathbf{a}_1,\mathbf{a}_2,...,\mathbf{a}_k,\mathbf{0}_{k+1},...,\mathbf{0}_n]\mathbf{S}$$

where $\hat{\mathbf{A}}=[\mathrm{a}_1,\mathrm{a}_2 \ldots \mathrm{a}_\mathrm{p}, 0_{\mathrm{p}+1}, 0_{\mathrm{p}+2} \ldots 0_\mathrm{n}]$. The crucial point of the above concept is proper A and S estimation. It is difficult task because we have not information which decomposition is most adequate. Therefore, we must test various transformations resulting in components with different properties. The most adequate methods to solve the first problem seem to be the blind signal separation (BSS) techniques.

Independent component analysis (ICA) is a statistical tool, which allows decomposition of observed variable X into independent components $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n]^T$ [3]. Typical algorithms for ICA explore higher order statistical dependencies in a dataset, so after ICA decomposition we have signals (variables) without any linear and non-linear statistical dependencies. To obtain independent components we explore the fact that the joint probability of independent variables can be factorized by the product of the marginal probabilities

(4) $$\overbrace{p_1(\mathbf{y}_1)p_2(\mathbf{y}_2)\dots p_n(\mathbf{y}_n)}^{q_\mathbf{y}(\mathbf{Y})} = \overbrace{p_{1\dots n}(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n)}^{p_\mathbf{y}(\mathbf{Y})}.$$

One of the most popular method to obtain (4) is to find such W that minimizes the Kullback-Leibler divergence between $p_\mathbf{y}(\mathbf{Y})$ and $q_\mathbf{y}(\mathbf{Y})$ [3]:

(5) $$\mathbf{W}_{opt} = \min_{\mathbf{W}} D_{KL}(p_\mathbf{y}(\mathbf{WX}) \,||\, q_\mathbf{y}(\mathbf{WX})) = \min_{\mathbf{W}} \int_{-\infty}^{+\infty} p_\mathbf{y}(\mathbf{Y}) \log \frac{p_\mathbf{y}(\mathbf{Y})}{q_\mathbf{y}(\mathbf{Y})} d\mathbf{Y}$$

There are many numerical algorithms estimating independent components like Natural Gradient, FOBI, JADE or FASTICA [1, 3].

Statistical analysis of noise

It seems intuitive and natural that for data with temporal structure the random noises are not regular or smooth. The standard characteristic investigated in this case is the autocorrelation function or its Fourier transformation called power spectrum [10, 11]. Unfortunately, it has some disadvantages like functional form what is difficult for comparison, it appears to be insensitive in some cases and it causes problems with detection of the long memory dependencies due to its exponential decrease. The alternative to autocorrelation function is the Hurst exponent H and R/S analysis [2, 6, 7]. It is important to calculate H only on linear part of the regression identified by individual inspection during analysis [3, 12].

Therefore, for signals with temporal structure we propose a following measure. Let us consider the signal y with temporal structure and observations indexed by k=1,2,…,N. The variability (and thus unpredictability of the signal) might be measured with the following formula:

(6) $$Q(y) = \frac{\frac{1}{N-1}\sum_{k=2}^{N} |y(k) - y(k-1)|}{\rho(\max(y) - \min(y))}$$

where symbol $\rho(u) = \begin{cases} u & \text{for } u \neq 0 \\ 1 & \text{for } u = 0 \end{cases}$

means Kronecker delta and it is introduced to avoid dividing by zero. The possible values of measure (6) are from 1 to 0. The measure has simple interpretation: it is maximal when the changes in each step are equal to range (maximal change), and is minimal when data are constant. In both cases the signal is totally predictable, but between those marginal states the signal is random.

To present some reference values let us calculate the value of Q for random signals from uniform and normal distribution. It can be easily demonstrated, by simulations, that for uniform distribution the noise factor is 1/3.

In case of a normal distribution of the expected value depends on the number of observations and is expressed by the following table (Table 1). See also Fig. 1.

Table 1. Expected range and variability of Q-measure for sample of random variable y~N[0,1]

No obs.	Expected range	Q
10	1.81	0.44
100	3.29	0.24
1000	4.37	0.18
10000	5.26	0.15

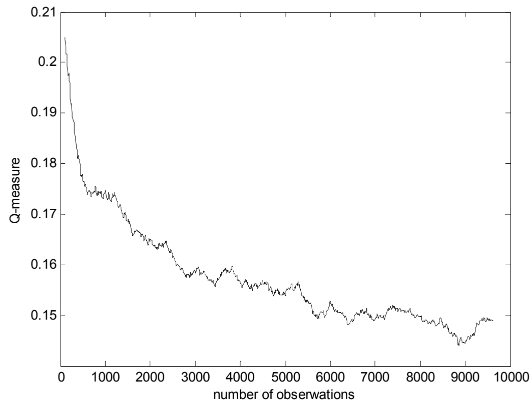

Fig. 1. Q-factor dependence on number of observations for Gaussian noise

The examples show that for limited random variables the Q-measure gives stable results, while for unlimited – it asymptotically equals 0.

To combine information about the distance between the reference noise and the Gaussian noise we propose the following noise factor P:

(7) $$P = \sum_{p \in K} k_p \left(Q_p(y) - Q_p(y_v)\right)^2$$

where k_p is the weight for each square difference to the reference noise; the default value is equal to one.

Numerical Experiment

The validation test of the proposed concept with noise detection was performed on the problem of load prediction in the Polish power system. Our task was to forecast the hourly energy consumption in next 24 hours based on the energy demand from last 24 hours and calendar variables: month, day of the month, day of the week, and holiday indicator. The data covered observations from 1988 until 1998.

We trained six MLP neural networks with one hidden layer (with 12, 18, 24, 27, 30, 33 neurons respectively). The quality of the results is measured with MAPE criterion for following neural networks M1:MLP12, M2:MLP18, M3:MLP24, M4:MLP27, M5:MLP31, M6:MLP33. For such primary models we perform their ensemble with BSS methods. Table 2 presents the results of primary models.

Table 2. Prediction results for primary models

Models	M1	M2	M3	M4	M5	M6
MAPE error	2.39	2.36	2.37	2.40	2.40	2.36

Prediction results of the primary models were then decomposed using ICA, please see Fig. 2.

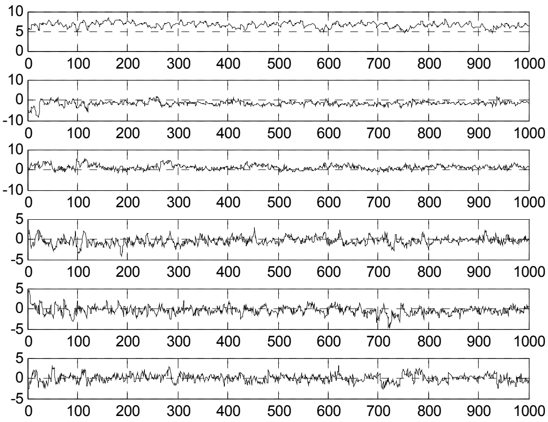

Fig. 2.Latent components after ICA

After decomposition we obtained six components for which we calculated the noise factor given in (7), see Table 3.

Table 3. The noise factor for each component after ICA

Component	y_1	y_2	y_3	y_4	y_5	y_6
Noise factor *P*	0.0959	0.0975	0.0799	0.0946	0.0871	0.0911

It is assumed that lower values of noise factor indicate similarity to the reference noise. Therefore, the rejection of them should improve the final results. And the other way round, we can easily identify the constructive components which are responsible for the core of prediction.

In our case we observed that the best prediction improvement is obtained after elimination of three components (y_4, y_5, y_6), not necessarily those with the lowest value. To benefit improvement rate, these components may be rejected individually or in combinations. The following table (Table 4) presents the smallest MAPE error after rejection of particular components or combination of them.

Table 4. The best models after rejection of particular components

Rejected component	MAPE error of the best model
y_4	2.32
y_5	2.28
y_6	2.24
y_4, y_5	2.33
y_4, y_6	2.29
y_5, y_6	2.25

Conclusions

In this article we mainly focused on the model's results decomposition based on independent component analysis in electric load prediction. We extended this concept with *a priori* approach for components identification based on novel formula for noise factor.

Our experiment on electricity data confirmed the validity of the proposed solutions. We could benefit of about 5-6% of MAPE reduction (best primary model vs. best model after decomposition). However, a number of research and methodological issues is still open. The most important include the way of proper identification and estimation of the reference noise.

The above presented identification method can be also addressed to wide area of data exploration models like simulations, trading systems, forecasting models or machine learning systems. Therefore, for the future research we see the need for the models that can identify the fundamental factors influencing, for instance, the stock market environment. Unfortunately, these factors are often hidden or mixed with noises. Consequently, a fundamental problem in financial market modeling is to estimate the main trends and to separate the general market dependencies from the individual behaviour of a given financial instrument.

Acknowledgements. The work was funded by the National Science Center based on decision number DEC-2011/03/B/HS4/05092.

REFERENCES

[1] Cichocki A., Amari S., Adaptive Blind Signal and Image Processing, John Wiley, Chichester, 2002

[2] Hurst H.E., Long term storage capacity of reservoirs, *Trans. Am. Soc. Civil Engineers*, No 116, 1951

[3] Hyvärinen A., Karhunen J., Oja E., Independent Component Analysis, John Wiley, 2001

[4] Misiorek A., Weron R., Modelowanie sezonowości a prognozowanie zapotrzebowania na energię elektryczną, *Energetyka,* Nr 12 (2004), 780-785

[5] Osowski S., Siwek K., Regularization of neural networks for improved load forecasting in the power system, *IEEE Proc. of Generation, Transmission and Distribution*, Vol. 149, No 3 (2002), 340-344

[6] Peters E., Fractal market analysis, John Wiley, 1996

[7] Samorodnitskij G., Taqqu M., Stable non-Gaussian random processes: stochastic models with infinitive variance. New York, London, Chapman and Hall, 1994

[8] Sprawozdanie CFTC „Staff report on commodity swap dealers and index traders with Commission recommendations", Waszyngton, 2008 r. American Economic Review; Komunikat Komisji COM(2008) 821 „Ceny żywności w Europie" oraz stanowiący uzupełnienie document roboczy służb Komisji SEC(2008) 2971 „Grupa zadaniowa ds. roli spekulacji w zmianach cen towarów rolnych — czy na rynkach towarowych rośnie bańka spekulacyjna?". http://www.loe.org/images/content/080919/cftcstaffreportonswapdealers09.pdf (01/08/2012)

[9] Szupiluk R., Wojewnik P., Zabkowski, T., Model Improvement by the Statistical Decomposition. Artificial Intelligence and Soft Computing. *Lecture Notes in Computer Science*, Vol. 3070(2004), 1199-1204

[10] Therrien C.W., Discrete Random Signals and Statistical Signal Processing, Prentice Hall, New Jersey, 1992, Toulouse, France, 2006

[11] Vaseghi S.V., Advanced signal processing and digital noise reduction, Chichester, John Wiley and Sons, Stuttgart, B. G. Teubner, 1997

[12] Yang Y., Adaptive regression by mixing, *Journal of American Statistical Association*, vol. 96, 2001

Authors: dr inż. Ryszard Szupiluk, Szkoła Główna Handlowa, 02-554 Warszawa, Al. Niepodległości 162, E-mail: rszupi@wp.pl; dr inż. Tomasz Ząbkowski, Szkoła Główna Gospodarstwa Wiejskiego, Wydział Zastosowań Informatyki i Matematyki, 02-787 Warszawa, ul. Nowoursynowska 159, E-mail: tomasz_zabkowski@sggw.pl