



# Improving Time Series Prediction via Modification of Dynamic Weighted Majority in Ensemble Learning

Marek Lóderer<sup>(✉)</sup>, Peter Pavlík, and Viera Rozinajová

Faculty of Informatics and Information Technologies,  
Slovak University of Technology, Ilkovičova 2, 841 04 Bratislava, Slovakia  
{marek\_loderer, xpavlikp, viera.rozinajova}@stuba.sk

**Abstract.** In this paper, we explore how the modified Dynamic Weighted Majority (DWM) method of ensemble learning can enhance time series prediction. DWM approach was originally introduced as a method to combine predictions of multiple classifiers. In our approach, we propose its modification to solve the regression problems which are based on using differing features to further improve the accuracy of the ensemble. The proposed method is then tested in the domain of energy consumption forecasting.

**Keywords:** Ensemble learning · Dynamic weighted majority method  
Time series prediction

## 1 Introduction

Time series predictions have indispensable importance for the human society and they have a great impact on many domains as well as on our everyday human activities. Predictions are used in many areas like industry, energetics, business, banking, weather forecasting, research, etc. Especially in energetics, accurate forecasts of the future values are crucial. Identifying the underlying patterns in the data is usually not a trivial task. In the last decades, researchers have introduced several prediction methods (predictors) to solve this problem [8].

Some of the predictors use mathematical and statistical calculations, like Linear Regression [9], ARIMA models or Exponential Smoothing [19]. Others are based on Machine learning, e.g. Support Vector Regression, Neural Networks and Random Forest [6, 7].

Each of these predictors can be successful at describing certain types of time series patterns but may become less accurate over time if the time series contains changes in concept characteristics known as Concept Drift. To solve this problem various learning methods such as Bagging [2], Boosting [17], Stacking [20] and numerous hybrid approaches [7, 10, 15, 21] have been developed.

The strength of ensembles lies in the fact that even if some of its predictors fail to predict the new pattern correctly, others with ability to predict accurately in

these changed conditions can compensate the overall prediction error. Therefore, highly diverse ensembles are effective in lowering the error after a concept drift occurs [13].

Another benefit of ensemble is the ability to adapt to changes by dynamically combining base members (predictors) according to their recent performance which increases the overall accuracy of the prediction [3].

Ensemble learning based on Dynamic Weighted Majority strongly employs dynamic combination of predictors along with predictors reweighting which is described and discussed in this paper.

This paper is organized as follows. In Sect. 2 we introduce the main concept of ensemble learning and related terms. In Sect. 3 we present our proposed ensemble learning model based on modification of Dynamic Weighted Majority method for time series prediction. In Sect. 4 we present results of experimental evaluation and our conclusions can be found in Sect. 5.

## 2 Ensemble Learning

In general, the main principle of ensemble learning is based on a proper combination of results of different base models (predictors or classifiers) that can create a more accurate result in comparison to the result provided by the best individual model [12].

Ensemble learning consists of three main subprocesses: ensemble generation, ensemble pruning and ensemble integration. In ensemble generation, a set of diverse base prediction models are trained. The required level of diversity of base models can be achieved by three main approaches - data, parameter and structural diversity [16].

Ensemble pruning is used to eliminate redundant and high erroneous models to increase the overall accuracy of final prediction. The pruning can be performed by various ranking, search or partitioning-based methods [12]. This part of the ensemble learning is optional.

The last subprocess, ensemble integration, provides combination of results of prediction models. The combination is usually carried out as a linear combination, where the weights are calculated by numerous approaches e.g. a simple mean, an inverse value of prediction model performance or more complex weighted schemes based on optimization algorithms [5, 21].

Several types of ensemble based on Outperformance method [1] or Dynamic Weighted Majority method [10] combine outputs of currently generated predictors taking into account values of past weights and predictors errors. This additional information helps ensemble to overcome high fluctuation of weights in noisy and quickly changing data.

### 2.1 Dynamic Weighted Majority

As mentioned earlier, Dynamic Weighted Majority (DWM) is an ensemble method which uses more complex weighted schemes. It was first described by

Kolter and Maloof in 2003 [10]. It is based on an older Weighted Majority algorithm from Littlestone and Warmuth which gives individual experts (prediction methods of the ensemble) weights, modifies them according to their performance and generates final prediction by combining the predictions of the experts with consideration to their weights [11].

While original Weighted Majority algorithm works with a static set of experts, the Dynamic Weighted Majority can add or remove a number of experts based on their performance. Thanks to this added feature, the ensemble can successfully predict even in a changing environment with occurrence of Concept Drift [10].

The original algorithm works with a set of experts with corresponding weights. Each iteration of the algorithm starts by calculating a global prediction of the current ensemble. The global prediction is obtained by combining predictions of all experts proportionally to their weights. The algorithm obtains a prediction from each member of the ensemble and adds its weight to the sum for the corresponding output class. The class with the highest weight is then set as the global prediction of the ensemble.

If the prediction does not match the sample label, the weights of incorrect experts are lowered by predefined multiplicative factor from interval (0,1). If a weight of any expert is lower than a predefined threshold value, then the expert is removed from the ensemble.

A new expert is trained and added every time the global prediction is incorrect. At the end of each iteration, the weights of the experts are normalized to add up to 1. Otherwise, the resulting prediction would be biased.

## 2.2 Modification for Regression

The Dynamic Weighted Majority was originally created for classification but the base idea of keeping a dynamic set of experts is applicable for regression as well. However, changes must be made in the process of evaluating experts' performance, modification of their weights and experts' replacement.

When solving classification problems, evaluating the correctness of the prediction is quite straightforward. On the other hand, a result of regression is a number from a continuous interval where the accuracy of prediction has to be measured by certain metrics. That means we cannot easily decrease the weight of an expert by a constant factor when its result is incorrect.

Subsequently, due to the property of regression problems another step of the algorithm cannot be directly used - adding a new expert when the global prediction is incorrect. A possible solution is setting a threshold to specify the highest acceptable error of a prediction method. But setting the threshold is very domain-specific and often undesirable. A better solution for general use is a constant size of the ensemble which means a new expert is added to the ensemble only in case when another expert has been removed.

In 2016, a paper *Prediction of Power Load Demand Using Modified Dynamic Weighted Majority Method* by Radoslav Nemeč et al. applied the Dynamic Weighted Majority on the regression problem of predicting power load demand

[14]. In this paper, the problem of reducing expert weights was solved by introducing an error constant  $\gamma$ . The error constant reduces the weights of experts who achieve higher error and increases weights of experts with more precise results. The problem of adding new experts was solved by an constant size of the ensemble.

### 3 Proposed Algorithm

In this paper, we propose a general version of Dynamic Weighted Majority for regression and time series prediction. Our method is based on the modification for regression, which was mentioned in the previous chapter, but without the need for error constant  $\gamma$ , as it is very dependent on data. After removing this constant, we can use the error of prediction as a measure to determine how much we want to reduce the weight of an expert, as opposed to binary choice of reducing or not reducing the weight by a given constant. We believe this approach can increase the accuracy of the ensemble.

---

**Algorithm 1.** Dynamic Weighted Majority for regression

---

```

 $\{\vec{x}, y\}_n^1$  : training data composed of attribute vector and class label
 $\beta$  : expert weight lowering factor,  $0 \leq \beta > 1$ 
 $\theta$  : threshold weight value for expert removal
 $p$  : period of expert replacement
 $\{e, w\}_m^1$  : set of experts and their weights
 $A, \lambda \in \{1, \dots, c\}$  : global and local predictions
 $m$  : size of the expert set in ensemble
 $\varepsilon_{1, \dots, m}$  : prediction errors
1:  $e_m \leftarrow CreateExpertSet()$ 
2:  $w_{1, \dots, m} \leftarrow 1/m$ 
3: for  $i \leftarrow 1, \dots, n$  do
4:    $A \leftarrow 0$ 
5:   for  $j \leftarrow 1, \dots, m$  do
6:      $\lambda \leftarrow Predict(e_j, \vec{x}_i)$ 
7:      $\varepsilon_j \leftarrow MAPE(\lambda, y_i)$ 
8:      $A \leftarrow A + \lambda * w_j$ 
9:    $w \leftarrow LowerWeights(w, \varepsilon)$ 
10:  if  $i \bmod p = 0$  then
11:    for  $j \leftarrow 1, \dots, m$  do
12:      if  $w_j < \theta$  then
13:         $e, w \leftarrow ReplaceWithNewExpert()$ 
14:   $w \leftarrow NormalizeWeights(w)$ 
15:  output  $A$ 

```

---

The algorithm of the proposed DWM method starts by creating a set of  $m$  different experts with equal weights (lines 1–2). An iteration starts by obtaining

a prediction of the training sample for all experts. These local predictions are multiplied by the weight  $w_j$  of an expert  $j$  and added to global prediction  $\Lambda$  (lines 3–8). Subsequently, the prediction errors are calculated for each expert and saved into a vector. The prediction error of each expert is calculated by Mean Absolute Percentage Error metric (1)

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|, \quad (1)$$

where  $n$  is the number of samples,  $A_t$  is the actual value and  $F_t$  is the predicted value.

Although, various error metrics can be used [18]. These accuracy values are transformed into a vector of multipliers from interval  $\langle \beta, 1 \rangle$  which is used to lower the weights of experts proportionally to their performance on the last sample (line 9).

In our implementation, we achieve this by using the transformation function (2) which assigns the lowest performing expert a multiplier of  $\beta$  and gradually increases the multipliers of other experts up to a theoretical maximum of 1 for perfect prediction:

$$\text{mult}_i = \left( 1 - \frac{\varepsilon_i}{100} \right)^{\frac{1}{\log_{\beta} \left( 1 - \frac{\max(\varepsilon)}{100} \right)}} \quad (2)$$

In case the expert replacement is allowed in this iteration (line 10), we check if any expert has weight lower than the threshold  $\theta$  (lines 11–12). If that is the case, expert is removed from the ensemble and replaced by a different one with an initial starting weight (line 13). At the end of the iteration, weights of the experts are normalized so the sum is equal to 1 (line 14).

## 4 Evaluation

We evaluated the accuracy of our proposed DWM method on time series data containing electricity consumption measurements. We predict electricity consumption for the next 24 h and then we move the prediction window to the next day.

For testing of the proposed ensemble, a process of creating diverse experts is needed. We fulfill this requirement by creating structurally diverse experts based on different prediction methods. A pool of experts is created from commonly used time series prediction methods, namely: Autoregressive Integrated Moving Average (ARIMA), Random Forest (RF), Feed-forward Neural Network with a single hidden layer and lagged inputs (NN) and Support Vector Regression (SVR). Each method is trained on a window of four weeks training data prior to the prediction date. However, we do not use these methods directly on the time series data.

Before prediction, the time series is split into seasonal, trend and remainder component by the *Seasonal and Trend decomposition using Loess (STL)* [4].

Each expert in ensemble consists of one seasonal, trend and reminder component, where each component can be computed by different prediction method. By this approach we can increase the number of possible experts up to  $n^3$  where  $n$  is the number of used base prediction methods.

The ensemble starts with a given number of random experts with equal weights. The weights are modified in each iteration by formula (2) and subsequently if a weight of any expert falls below the threshold  $\theta$ , then it is replaced by another expert from currently unincluded experts. In our implementation, we randomly pick one of the experts included in the ensemble and mutate it by changing one of its three components to create a new expert.

## 4.1 Data

The proposed ensemble model was evaluated on two electricity consumption datasets. Both datasets contain energy measurements from households as well as from enterprises. The first dataset consists of time series measurements with 60 min period from Toronto region, Canada. The data are collected by the Independent Electricity System Operator<sup>1</sup>. In our experiments, we used the sliding window approach to perform daily predictions for whole year 2011.

The second dataset consists of time series measurements with 30 min period from Australian Energy Market Operator<sup>2</sup>. In the experiment, we used aggregated data from the state Tasmania. Daily predictions were computed on data from year 2009.

## 4.2 Results

In our experiments we tested prediction accuracy of the proposed ensemble. Since the ensemble has several configuration parameters that strongly affects the prediction outcome, at first we experimentally estimated optimal values for these parameters. The parameter  $\beta$  was set to 0.65, parameter  $\theta$  to 0.5 and  $p$  was 1.

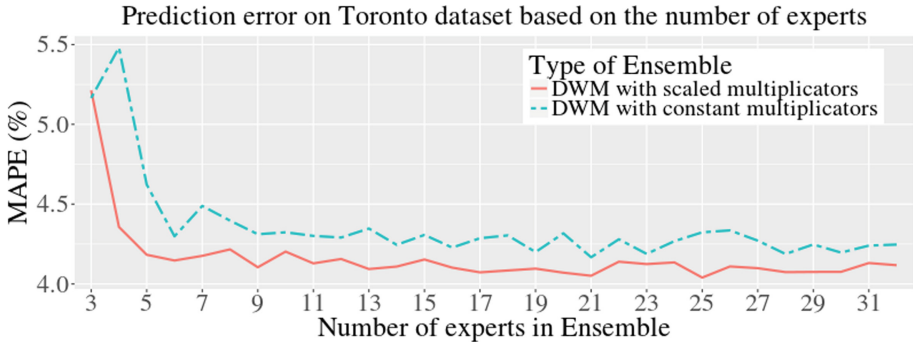
The parameter estimation was calculated on one whole year of previous measurements in Toronto (year 2010) and Tasmania (year 2008) datasets. This one-year period of previous data was also used to eliminate potential prediction error caused by randomness of initial experts in the ensemble, and to select appropriate ones. Another important aspect influencing prediction accuracy of the ensemble is the number of experts in it.

To put the results of the proposed DWM ensemble with scaled multipliers into perspective, we also measured accuracy of the DWM ensemble with constant weights multipliers based on the work of Radoslav Nemeč et al. [14].

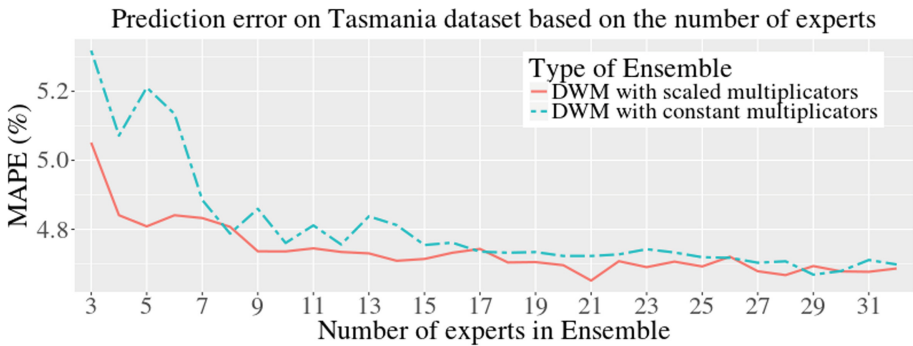
In our first experiment we evaluated the prediction accuracy based on the number of experts in the ensemble. An error metric MAPE was used to evaluate the prediction accuracy. Figures 1 and 2 show the development of the average

<sup>1</sup> <http://www.ieso.ca/>.

<sup>2</sup> <http://www.aemo.com.au>.



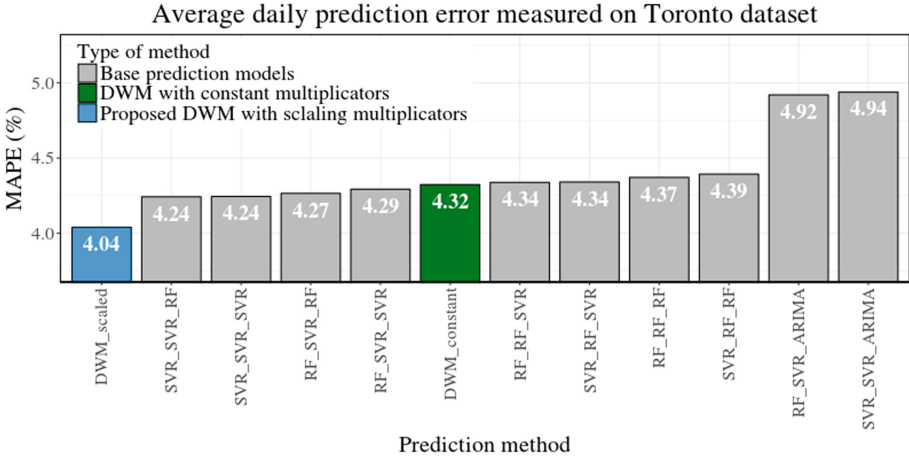
**Fig. 1.** Accuracy comparison of DWM ensembles with scaled and constant multipliers based on a number of experts in the ensemble measured on the electricity consumption dataset from Toronto region in Canada.



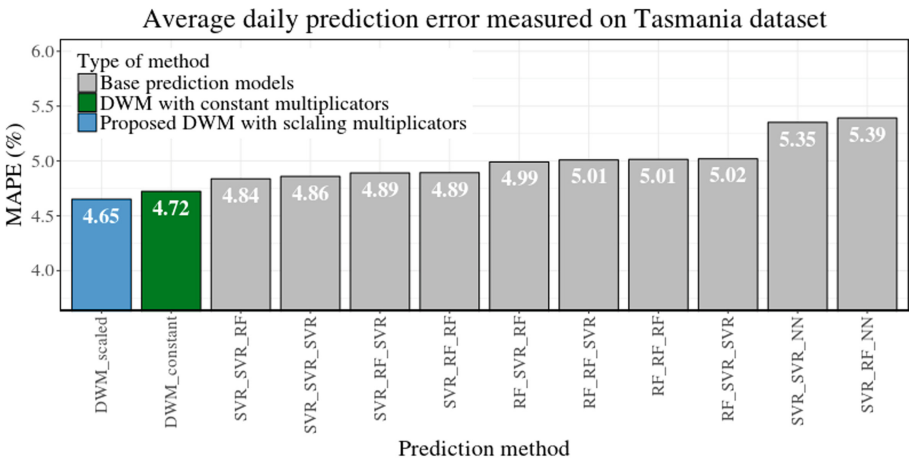
**Fig. 2.** Development of prediction accuracy of DWM ensembles with scaled and constant multipliers based on a number of experts in the ensemble evaluated on the Tasmania dataset.

daily prediction error of tested ensembles based on different number of experts. The results show that a reasonable number of experts in the ensemble is about 5 to 9. A higher number of experts improves results only slightly and it increases the computational complexity. The results also show that the proposed DWM ensemble with scaled multipliers obtained lower prediction error in comparison to the DWM ensemble with constant weights multipliers in almost all tested cases.

The second experiment was designed to show prediction accuracy of the tested DWM ensembles and 10 best experts. Results displayed in Figs. 3 and 4 show average daily prediction error measured on Toronto dataset for time period from 1.1.2011 to 31.12.2011 and Tasmania dataset from 1.1.2009 to 31.12.2009. In Toronto dataset, the number of experts in ensemble was set to 25. In case of Tasmania dataset, we used 21 experts.



**Fig. 3.** Comparison of prediction error of DWM ensembles and ten best prediction methods (experts) on Toronto dataset in year 2011 displayed in ascending order.



**Fig. 4.** Results representing prediction error of tested DWM ensembles and their best performing member methods (experts) on Tasmania dataset in year 2009 displayed in ascending order.

As mentioned previously, each expert is composed of seasonal, trend and reminder component of the time series that is predicted by individual prediction method and combined to create final prediction for the next day. The name of an expert consists of abbreviations of used prediction methods. The position of abbreviation in the expert name represents the seasonal, trend and reminder component.

According to the results the proposed DWM ensemble with scaled multipliers outperformed DWM ensemble with constant multipliers as well as



best expert on both datasets. In Tasmania dataset, several experts obtained even better prediction results than DWM ensemble with constant multipliers.

The results also show that the majority of the seasonal, trend and reminder components of the best predicting experts in both testcases were predicted mainly by Support Vector Regression and Random Forest.

## 5 Conclusion

Our proposed modification of Dynamic Weighted Majority with scaling multipliers for regression and time series prediction has proven to be a successful approach to combine multiple predictors (experts) into an accurate ensemble. It is especially useful if many predictors with various accuracies are available, as it is able to identify the best performing predictors and omit the underperforming ones even in a changing environment.

We also compared our proposed ensemble with another DWM ensemble on two publicly available electricity consumption datasets. According to the results our solution outperformed all base prediction methods as well as other tested ensemble in terms of prediction accuracy.

However, there is still room for improvement. The possible aim of future research is to find an optimal transformation of forecast errors into multipliers of experts' weights. Replacing the exponential scaling of errors shown above by other mathematical transformations would undoubtedly impact the precision of the ensemble and should be explored further. Another direction of future research could involve optimization of constant values  $\beta$  and  $\theta$  used as the parameters of the ensemble.

**Acknowledgment.** This work was partially supported by the Slovak Research and Development Agency under the contract APVV-16-0213.

## References

1. Adhikari, R., Agrawal, R.K.: Performance evaluation of weights selection schemes for linear combination of multiple forecasts. *Artif. Intell. Rev.* **42**(4), 529–548 (2012). <https://doi.org/10.1007/s10462-012-9361-z>
2. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**(2), 123–140 (1996)
3. Cerqueira, V., Torgo, L., Oliveira, M., Pfahringer, B.: Dynamic and heterogeneous ensembles for time series forecasting. In: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 242–251, October 2017. <https://doi.org/10.1109/DSAA.2017.26>
4. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: STL: a seasonal-trend decomposition procedure based on loess. *J. Off. Stat.* **6**(1), 3–73 (1990)
5. Ezzeddine, A.B., et al.: Using biologically inspired computing to effectively improve prediction models. *Int. J. Hybrid Intell. Syst.* **13**(2), 99–112 (2016)
6. Górriz, J.M., Puntónet, C.G., Salmerón, M., de la Rosa, J.J.G.: A new model for time-series forecasting using radial basis functions and exogenous data. *Neural Comput. Appl.* **13**(2), 101–111 (2004). <https://doi.org/10.1007/s00521-004-0412-5>

7. Halaš, P., Lóderer, M., Rozinajová, V.: Prediction of electricity consumption using biologically inspired algorithms. In: 2017 IEEE 14th International Scientific Conference on Informatics, pp. 98–103, November 2017
8. Hyndman, R.J., Athanasopoulos, G.: *Forecasting: Principles and Practice*. OTexts (2014)
9. Kantikoon, V., Kinnares, V.: The estimation of electrical energy consumption in abnormal automatic meter reading system using multiple linear regression. In: IEEE International Conference on Electrical Machines and Systems (ICEMS), pp. 826–830 (2013)
10. Kolter, J.Z., Maloof, M.A.: Dynamic weighted majority: an ensemble method for drifting concepts. *J. Mach. Learn. Res.* **8**(Dec), 2755–2790 (2007)
11. Littlestone, N., Warmuth, M.: The weighted majority algorithm. *Inf. Comput.* **108**(2), 212–261 (1994)
12. Mendes-Moreira, J., Soares, C., Jorge, A.M., Sousa, J.F.D.: Ensemble approaches for regression. *ACM Comput. Surv.* **45**(1), 1–40 (2012)
13. Minku, L.L., White, A.P., Yao, X.: The impact of diversity on online ensemble learning in the presence of concept drift. *IEEE Trans. Knowl. Data Eng.* **22**(5), 730–742 (2010)
14. Nemeč, R., Rozinajová, V., Lóderer, M.: Prediction of power load demand using modified dynamic weighted majority method. In: Świątek, J., Tomczak, J.M. (eds.) *ICSS 2016. AISC*, vol. 539, pp. 36–49. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-48944-5\\_4](https://doi.org/10.1007/978-3-319-48944-5_4)
15. Qiu, X., Zhang, L., Ren, Y., Suganthan, P.N., Amaratunga, G.: Ensemble deep learning for regression and time series forecasting. In: 2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL), pp. 1–6, December 2014. <https://doi.org/10.1109/CIEL.2014.7015739>
16. Ren, Y., Zhang, L., Suganthan, P.N.: Ensemble classification and regression—recent developments, applications and future directions [review article]. *IEEE Comput. Intell. Mag.* **11**, 41–53 (2016)
17. Schapire, R.E.: The boosting approach to machine learning: an overview. In: Denison, D.D., Hansen, M.H., Holmes, C.C., Mallick, B., Yu, B. (eds.) *Nonlinear Estimation and Classification*. LNS, vol. 171, pp. 149–171. Springer, New York (2003). [https://doi.org/10.1007/978-0-387-21579-2\\_9](https://doi.org/10.1007/978-0-387-21579-2_9)
18. Shcherbakov, M., Brebels, A., Shcherbakova, N., Tyukov, A., Janovsky, T., Kamaev, V.: A survey of forecast error measures. *World Appl. Sci. J.* **24**, 171–176 (2013)
19. Taylor, J.W., McSharry, P.E.: Short-term load forecasting methods: an evaluation based on european data. *IEEE Trans. Power Syst.* **22**(4), 2213–2219 (2007)
20. Ting, K.M., Witten, I.H.: Issues in stacked generalization. *CoRR abs/1105.5466* (2011). <http://arxiv.org/abs/1105.5466>
21. Xiao, L., Wang, J., Hou, R., Wu, J.: A combined model based on data pre-analysis and weight coefficients optimization for electrical load forecasting. *Energy* **82**, 524–549 (2015)