International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

GEFCom 2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach

Gábor I. Nagy [a,c,d,*], Gergő Barta [a,c], Sándor Kazi [a,c], Gyula Borbély [b], Gábor Simon [c]

[a] Budapest University of Technology and Economics, Hungary
[b] Ericsson, Hungary
[c] Dmlab, Hungary
[d] Enbritely, United Kingdom

ARTICLE INFO

Keywords:
Probabilistic forecast
Ensemble methods
Renewable energy
Quantile regression
Gradient boosting regression
Quantile regression forest

ABSTRACT

We investigate the probabilistic forecasting of solar and wind power generation in connection with the Global Energy Forecasting Competition 2014. We use a voted ensemble of a quantile regression forest model and a stacked random forest – gradient boosting decision tree model to predict the probability distribution. The raw probabilities thus obtained need to be post-processed using isotonic regression in order to conform to the monotonic-increase attribute of probability distributions. The results show a great performance in terms of the weighted pinball loss, with the model achieving second place on the final competition leaderboard.

1. Introduction

Renewable energy is becoming increasingly important in satisfying world energy demands. One particular characteristic of renewable energy is that the output of a power plant depends largely on weather conditions, and accurate forecasts of power generation are an inevitable requirement in order to integrate it into the power grid efficiently. The Global Energy Forecasting Competition 2014 featured a unique approach to the forecasting of hourly power outputs: the forecasting of a probability distribution of the target variable, rather than forecasting one single value (Hong et al., 2016). This methodological difference from single value forecasts is huge, because it provides stakeholders in the industry with more information to incorporate into their daily work. As a side effect, new methods must be used both efficiently and accurately for producing probabilistic forecasts. After useful methods for the task had been assessed, our team's efforts focused on providing an ensemble of probabilistic forecasts from a number of different methods, such as gradient boosting, a method that had been successful in many previous competitions, including the AMS solar power forecasting competition, and quantile regression forests, which offer a non-linear, variable interaction sensitive method that can predict probability distributions directly (Aggarwal & Saini, 2014). In this paper, we show how the blend of these two methods can achieve good results in the field of renewable energy, in terms of the pinball loss.

2. Modeling methods

Previous work has shown that using multiple regressors is often better than using only one (Rokach, 2010). Therefore, we employed regressor ensembling in our approach in a number of ways from the beginning of the competition. In general, we used four types of ensembling:

* Corresponding author at: Budapest University of Technology and Economics, Hungary.
E-mail address: nagyg@tmit.bme.hu (G.I. Nagy).

http://dx.doi.org/10.1016/j.ijforecast.2015.11.013
0169-2070/© 2016 Published by Elsevier B.V. on behalf of International Institute of Forecasters.

voting, bagging, boosting and stacking predictors. Voting was found to be particularly useful for averaging the quantile forecasts of different models.

Both standard random forests and quantile regression forests implement bagging. The random forest (RF) is perhaps the best-known ensemble method; it combines decision trees to achieve an improved predictive performance, as was introduced by Breiman (2001). In addition to being a great out-of-the-box model, it also offers various useful features: it provides an intrinsic evaluation of the results based on the data discarded by bootstrapping (called the out-of-bag error), and variable importance estimates are also provided. One enhancement of random forests is the quantile regression forest (QRF) method (Meinshausen, 2006). It provides a non-parametric way of estimating conditional quantiles instead of the conditional mean, while the power of the high-dimensional regression of random forests can be exploited at the same time. To build a quantile regression forest model, one should grow a number of trees, as when using a random forest regressor, but take note of all observations for each leaf of each tree, not just the average. QRF has the beneficial property that the target quantile does not need to be set in the training phase, but only during model application, which reduces the training time significantly if a lot of quantiles are defined.

Boosting was used in gradient boosting decision trees (GBDT), a model that has been used successfully in various other fields, as well as in GEFCOM 2012 (Friedman, 2002). The predictor generated in gradient boosting is a combination of weak decision tree learners, which were built iteratively using the negative gradient of a loss function. The final predictor will be the weighted combination of these predictors. There are various benefits of utilizing boosting: various risk functions are applicable, and intrinsic variable selection is carried out. It also resolves multicollinearity issues and works well with large numbers of features without overfitting. In addition, GBDTs have the ability to learn quantile loss functions directly. This means that one can train k different models for different quantiles in order to acquire the probability distribution at the desired granularity; in our case, k was set to 99. The raw probability distribution generated by this method does not necessarily increase monotonically, so isotonic regression, a noise-reduction technique used to transform a series of observations into a non-decreasing stepwise function, was applied to ensure that all predictions conform to the law of cumulative probability distributions (de Leeuw, Hornik, & Mair, 2009).

Stacking was used in both tracks: a random forest was trained using 10-fold validation and a least squares loss function. The power output estimates of this simple point-forecast random forest (RF) were then used as features in the final GBDT modeling (not to be confused with the QRF).

The following sections describe the unique data processing and modeling approach used in each track.

3. Solar track

3.1. Data description and preparation

The task for the solar track was to predict 99 quantiles of the normalized power outputs of three different solar farms, referred to as zones. The four accumulated variables were used to derive five regular variables: one standard decumulated variable for each zone, and a precipitation-specific decumulated variable (TP), where the predictions were considered only for the possible non-zero output time ranges. Time-related information was also extracted from the given timestamp: the hour in two different formats, day of the year, and month. These generated variables (and four time related features) were the only additional variables used during the competition (see Table 1).

Table 1
Most important derived solar features and their description.

Variable name	Description
TCLW	Total column liquid water (kg m^{-2})
TCIW	Total column ice water (kg m^{-2})
SP	Surface pressure
RH	Relative humidity at 1000 mbar (%)
TCC	Total cloud cover
10U	10 m U wind component
10V	10 m V wind component
2T	2 m temperature
SSRD[_UNCUM]	Surface solar rad down (de)cumulated
STRD[_UNCUM]	Surface thermal rad down (de)cumulated
TSR[_UNCUM]	Top net solar rad (de)cumulated
TP[_UNCUM]	Total precipitation (de)cumulated
HOUR	The hour extracted from the timestamp as an integer value
HOUR_STRING	Hour extracted from timestamp as a discrete (string) value
DAY_OF_YEAR	The day of the year, calculated from the timestamp
MONTH	Month of measurement

3.2. Model updates

There were two significant changes to the methodology used in the GBDT model. The QRF model remained the same throughout the competition, apart from introducing the hour as a factor variable in Task 7. Task 8 introduced the decumulated TP variable in GBDT, and began to replace forecasts in the 11–18 h time range with 0.0. Task 9 introduced an RF–GBDT stacked modeling framework. From Task 10, voting the QRF and the RF–GBDT stack was abandoned on Zone 1, and only the stacked model was used, because the error distribution of QRF had a much higher mode, and would have caused a deterioration in the ensemble performance.

3.3. Performance evaluation

The dataset was split by day: the first and second days were sent into the training set sequentially, and the third day was sent into the validation set. This validation scheme was employed in single model performance testing. The ensemble performance was tested after a task had ended and the targets were published. The change in the trimmed mean error that was observed after the modeling tweaks is relatively small: Tasks 4–8 resulted in 0.1241, and this reduced to 0.1204 in Tasks 9–15. However, this small change meant a rather significant shift on the leaderboard (it being a very close competition). Fig. 1 shows the