International Journal of Forecasting 35 (2019) 1387–1388

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Energy forecasting in the big data world

Modern information and communication technologies have brought big data to virtually every segment of the energy and utility industries. While forecasting is an important and necessary step in the data-driven decision-making process, the problem of generating better forecasts in the world of big data is an emerging issue and a challenge to both industry and academia. This special section aims to collect top-quality forecasting articles that document cutting-edge research findings and best practices on a wide range of important business problems in the energy industry. Our emphasis is on big data, such as forecasting with high resolution data, the use of high-dimensional processes, forecasting in real-time, and the use of non-traditional data and variables.

As part of the preparation for this special section, we organized two events, the Global Energy Forecasting Competition 2017 (GEFCom2017) and the 2017 International Symposium on Energy Analytics (ISEA2017), in order to gather forecasting experts and practitioners to share their experience and insights. Through a thorough peer review process, we collected 14 papers for this special section. Eight of them are from GEFCom2017, while the other six are from regular submissions covering diverse topics in the areas of energy supply, demand and price forecasting.

GEFCom2017 papers

GEFCom2017 is the third and most challenging of the Global Energy Forecasting Competition series. The theme was hierarchical probabilistic load forecasting, merging the challenges of GEFCom2012 and GEFCom2014. The competition included both a qualifying match and a final match. The qualifying match presented a problem at a modest scale, with the aim of educating the crowd and advancing the promising methods to the final. The final match required participants to forecast the electric load at the medium/low voltage level for over 100 delivery point meters. Hong, Xie, and Black introduce the competition setup and data, summarize the top-ranked methods, present several reflections on the competition series, and set a vision for future energy forecasting competitions.

We collected five papers from the qualifying match. Ziel applied a simple but effective method based on quantile regression. While all other winning teams more or less adopted some machine learning techniques, Ziel, under the team name "Simple but Good", was the only team to use a pure statistical method. Dimoulkas, Mazidi, and Herre developed their models based on neural networks. Two components of their method appeared to help improve the quantile score significantly, namely a stepwise feature selection procedure and a dynamic temperature scenario selection procedure. Smyl and Hua used three machine learning techniques: gradient boosting, quantile random forest, and neural networks. The two team members took separate data preprocessing approaches. Their final forecast is the ensemble of the forecasts from two branches. Landgraf, as a single-person team, also took an ensemble approach. The techniques he used included quantile regression, quantile regression forest, and gradient boosting machine. Making the probabilistic forecasts coherent across the hierarchy was not a requirement in GEFCom2017, and most of the teams did not try to take advantage of the hierarchy either. However, Roach made an effort to adjust the forecasts to ensure that the bottom-level zonal forecasts summed correctly to the aggregated zonal forecasts, and also showed that such a reconciliation improves the forecast accuracy.

The final match presented a much more advanced problem than the one in the qualifying match. Even visualizing the hourly demand series from 100+ meters can be a challenge. de Hoog and Abdulla introduced several visualization techniques to help in understanding the data quality issues, the relationships between different entities in the dataset, and the relevance of custom date ranges. The No. 1 team in the final match consisted of two meteorologists from Japan, Kanda and Quintana Veguillas. Compared with the other winning teams, they performed some unique treatments of the weather data, such as coarse-graining the relative humidity values, and eliminating any weather stations with many periods of flat temperature profiles.

https://doi.org/10.1016/j.ijforecast.2019.05.004

Non–GEFCom2017 papers

The GEFCom2017 papers present the practical methods and ideas of the top-ranked teams, but the six papers collected from the regular submissions also provide insightful research findings. Haben, Giasemidis, Ziel, and Arora analyzed several methods for both point and probabilistic load forecasting at 100 low voltage feeders, which are comparable to the ones in the final match of GEFCom2017. Some of their findings confirm the existing results in the literature, such as the feeder size being the biggest determining factor for forecast accuracy, while others of their findings are novel, and still others are controversial. One example of a novel finding is that technologies have significant implications for the forecast accuracy, while a controversial finding was that temperature is not an important factor in driving the forecast accuracy, though they emphasized that their conclusion is based only on a small area in the UK for a two-month period in winter. This also pushes for more extensive empirical studies in the future.

Most wind power forecasting studies in the literature test the proposed methodologies on one or a few sites. As more and more wind turbines are installed worldwide, a wealth of data is becoming available to enable advanced models to improve their very short term forecasts. Messner and Pinson proposed a time-adaptive lasso estimator and an efficient coordinate descent algorithm for updating the parameters of vector autoregressive models recursively online. They tested the proposed idea on wind power data from 172 sites in Western France and 100 wind farms in Denmark.

Solar power forecasting was recognized as an immature field during GEFCom2014. Despite the progress that has been made in the recent decade, most algorithms that are proposed in the academic literature are not practical enough to be used by independent system operators (ISOs). One reason for this is that solar forecasting research often overlooks or downplays time-related parameters in the production environment, such as forecast resolution and lead time. Yang, Wu, and Kleissl proposed an operational solar forecasting algorithm that is aligned closely with the real-time market forecasting requirements of the California ISO (CAISO). The proposed algorithm was evaluated over a two-year period at seven locations in five climate zones.

Marcjasz, Uniejewski, and Weron contributed two electricity price forecasting articles to this special section. One is on day-ahead markets, where they studied the importance of the long-term seasonal component. While this component is often ignored in short term electricity price forecasting, they showed that including it in the model led to better forecasts. The other paper is on the intraday market, where they applied the least absolute shrinkage and selection operator (LASSO) to gain insights on variable selection for very short-term electricity price forecasting. The recommendations that they made were based on the performances of 12 models over a set of prices in the German EPEX market. Their study should be beneficial to price forecasters who are trying to understand intraday market dynamics.

The last paper that we collected is on crude oil price forecasting. Li, Shang, and Wang proposed a method, based on online media text mining, for capturing the immediate market antecedents of price fluctuations. This is one of the first attempts to use deep learning techniques to extract hidden patterns within online news media for crude oil price forecasting. The empirical results favored the proposed topic-sentiment synthesis forecasting models over the benchmarks. Moreover, the forecasts can be improved by combining the text features and financial features.

Final remarks

The original call for papers included several important topics that have not been covered by this special section. For instance, we are interested in forecasting failures at both the component level and the system level, such as failures of wind turbines and transformers, and interruptions of power delivery service. Since data quality plays a vital role in the forecast accuracy, we are interested in novel solutions to existing data quality problems, as well as in practical methods of dealing with emerging data quality issues. In the big data era, the energy industry is becoming more and more customer-centric, and we are interested in applying forecasting techniques to help energy companies better understand and communicate with their customers. All of these areas need attention from both researchers and practitioners.

This is the third and final energy forecasting special section that we have edited in the 2010s for the *International Journal of Forecasting*. We first want to thank the former Editor-in-Chief Rob J Hyndman for starting this initiative in energy and helping with the Global Energy Forecasting Competitions. Many authors who contributed to this special section gathered and shared ideas at ISEA2017, which was hosted by Rob Hyndman, George Athanasopoulos, Pam Stroud, and the local organization team of the 2017 International Symposium on Forecasting, so a big thank-you to them. We also want to thank the interim Editor-in-Chief Esther Ruiz for her diligent work to ensure the timely publication of the issue. Last but not least, we want to thank all of the authors and reviewers who were involved in the peer review process. They helped improve the quality of all of the papers in this special section. We hope that readers will enjoy reading this special collection of energy forecasting papers!

Tao Hong *
Systems Engineering and Engineering Management, University of North Carolina at Charlotte, United States
E-mail address: hongtao01@gmail.com.

Pierre Pinson
Centre for Electric Power and Energy (CEE), Technical University of Denmark, Denmark

* Corresponding editor.