Post-doc: Aggregation of Experts for Out-of-Data Forecasting

Yannig Goude*, Olivier Wintenberger†

Position description

The candidate should start in September 2021. The position will be held at Sorbonne University and will long for 1 year. The salary is approximately 60 000 euros gross per year. The position is supported by the ANR T-REX and EDF R&D. The candidate will have to implement and apply the algorithm in R or Python on electricity data.

State of the art: Robust Online Aggregation of Experts

Initiated by Bates and Granger (1969), combination of forecasts is a powerful tool for predictive learning tasks (see e.g. Bennett et al. (2007)). On the last decade, ensemble methods and aggregation of experts have proven to be very efficient methods in the context of online time series forecasting. It was applied to very different real world contexts such as, among others: finance (Amat et al. (2018)), energy (Devaine et al. (2013), Gaillard et al. (2016), Nowotarski and Weron (2018), Uniejewski and Weron (2018)), meteo (Taillardat et al. (2016), Thorey et al. (2017)) and pollution forecasting (see Debry and Mallet (2014), Baudin (2015), Auder et al. (2016)). Many arguments are in favor of aggregation of experts: the existence of different models designed to forecast a similar quantity (e.g. the weather forecasting models in used in the different weather institute in the world), the development of machine learning tools proposing more and more predictive algorithms and the increase of data sets from different sources (sensors, iot, web...). As stated in Breiman (2001) averaging models can lead to variance reduction inducing better generalization errors and a key point is the diversity of the experts considered in the aggregation (see Brown et al. (2005)). Coupled with robust online aggregation algorithms (see Wintenberger (2017)) this leads to good forecasting performances even in adversarial environments.

Usually no hypothesis are set on the data generative process, allowing experts coming from very diverse methods (different statistical setting, physical models, expert advice...). To forecast y_{n+1} according to its past values $y_1, \dots, y_n$, we suppose to have access to a set of N experts producing forecasts of the sequence at each instant t -usually these experts are also based on some exogenous information X also observed online-. After that, aggregation is computed step by step: $\hat{y}_t = \sum_{j=1}^{N} \hat{p}_{j,t} \hat{y}_t^j$ where the weights are updated according to past performances of each experts. Many algorithms have been derived for that purpose since the seminal works of Littlestone and Warmuth (1994), Freund and Schapire (1997) and Vovk (1998), among others, one can cite Herbster and Warmuth (1998) where the fixed-share algorithm is proposed to track the best expert changing with time, Devaine et al. (2013) extend it to the case of specialized experts (experts activate during periods for which they are supposed to be designed), Gaillard et al. (2014);

*Chef de projet Energy Data Lab Edf R&D, `yannig.goude@edf.fr`
†Professeur LPSM, Sorbonne University, `olivier.wintenberger@upmc.fr`

Wintenberger (2017) proposed online aggregation algorithms fully automatized in the sens that the multiple learning rates are sequentially updated using past data. De Rooij et al. (2014) propose the FlipFlop algorithm, a sequential prediction algorithm that performs well when the data generation process is non-adversarial which is useful for many real data applications.

Description of the project: Online learning and Expert Design for Out-of-Data Forecasting

Guided by real case application and following the idea that sequential prediction algorithm can be improved in a non-adversarial setting, we proposed recently (see Adjakossa et al. (2020)) new algorithms and theoretical guarantees in the case where the experts come from Kalman recursions, fitting state-space models. We showed the interest of the approach for electricity consumption forecasting and we believe that this technique could be useful in a lot of field where post processing of expert forecasts by the mean of state space model could be done (e.g in the field of meteorological forecasting Zamo (2016) extending state-space models to probability forecast).

Online aggregation of expert is classically based on convex aggregation of a given set of experts "a-priori" given. To obtain good performances of the aggregation, it is recommanded to aggregate "diverse" experts, coming for various methods and/or using different sources of information (see Gaillard and Goude (2014)). In ensemble methods frameworks (random forest, boosting), it is well known that increasing the diversity of the base learners improve the ensemble performance. This can be done implicitly using boostrap as in the Bagging or explicitly by optimising some diversity criteria to enforce base learners to be diverse Brown et al. (2005); Reeve and Brown (2018); Bourel et al. (2020). In practice some improvements are made by Adjakossa et al. (2020) compared to previous aggregation of experts since dynamical Kalman recursions have a random walk behaviour, exploring the space in different directions. More precisely, the experts comes from a class of process such as in Gourieroux and Robert (2006) that can be stable for some periods and explosive for others in a random way.

This success for out-of-data prediction motivates the project since abrupt changes in dynamics have been observed recently due to the pandemic situation. We will focus here on improving forecasts in explosive and unstable periods. Our practical motivation is driven by improving energy forecasting (renewable production or electricity consumption) during very difficult periods such as extreme weather condition (cold and heat waves, atypical interseason). Another source of unstability is churn (losses and gains of customers) in the case of electricity consumption or the development of renewables production and self-consumption.

We propose to study aggregation for out-of-data prediction both in practice and in theory. In order to do so, we believe that one has to extend the previous works to experts whose one controls/quantifies their exploration of the space in a precise way. A way to generate diversity is in the choice of the loss function. First we will study how to obtain diverse experts using quantile/expectile losses. One way to do so is more precisely to consider the case where the experts are driven by a Generalized Extreme Value. For both kind of experts we will study their prediction ability, i.e. accuracy and uncertainty. Then we will use the uncertainty associated with each expert in the aggregation procedure. Doing so, we will focus on both aspect of expert generations and associated online aggregation algorithm in the context of unstable forecasting. Another option is to extend experts extreme predictions using some subsampling strategies. For instance generalized random forests can already produce out-of-data quantiles predictions.

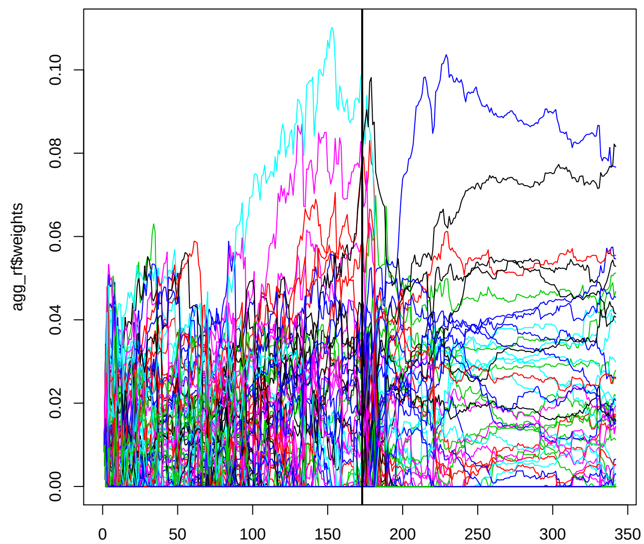

Figure 1: Online weights attributed to 100 trees of a random forest in function of time. The random forest is trained to forecast French electricity load during a normal period, a break (vertical black line) occurs due to the COVID19 pandemic affecting the weight distribution.

Such expert predictive extrapolation depends heavily on which level of the extreme we tuned the expert. Therefore the aggregation also depends on how the experts were tuned. The novelty of the project is to go in the opposite direction, using the aggregation score for tuning the experts. More precisely, if some expert predictive extrapolation at a high level is loosing confidence weight in the aggregation strategy, relying on this information one should change the extreme level used for tuning the prediction. The approach will be summarized such as a change of measure; the experts are predicting levels from a measure provided by their model. In a non-stationary setting, the prediction will deviate from this distribution. The confidence weights of the aggregation will be used as a proxy for this change of measure. When this change of measure is significative, one will use this information for tuning accordingly the extreme level of the experts.

Applications and expected outcomes

We will consider 2 real data sets in this project:

- RTE data of electrical consumption (see `https://www.rte-france.com/fr/eco2mix/`). It consists in an open data set of electrical data from 2012 to now at an half-hourly resolution.
- EDF data. A private data set corresponding to total electricity consumption of EDF customers from 2014 to now at an half-hourly resolution (subject to a confidentiality agreement).

In addition to these electricity data sets we will use meteorological data from Meteo France (meteo stations in France including temperature, solar radiation and wind speed). On these two datasets our goal is to improve forecasts at intraday (from 30 minutes to 24 hours) and daily horizons, focusing on special periods such as cold and heat waves more generally atypical periods.

The expected outcomes are potentially 2 publications, one in a computer science conference and one in a statistical learning/times series forecasting journal. It is also expected that the candidate takes part in the different events organized by the ANR T-REX.

References

Adjakossa, E., Y. Goude, and O. Wintenberger (2020). Kalman recursions aggregated online. *arXiv preprint arXiv:2002.12173*.

Amat, C., T. Michalski, and G. Stoltz (2018). Fundamentals and exchange rate forecastability with simple machine learning methods. *Journal of International Money and Finance 88*, 1–24.

Auder, B., M. Bobbia, J.-M. Poggi, and B. Portier (2016). Sequential aggregation of heterogeneous experts for pm10 forecasting. *Atmospheric Pollution Research 7*(6), 1101–1109.

Bates, J. M. and C. W. Granger (1969). The combination of forecasts. *Journal of the Operational Research Society 20*(4), 451–468.

Baudin, P. (2015, November). *Sequential prediction with ensemble aggregation: application to meteorological prediction with uncertainties* . Theses, Université Paris 11.

Bennett, J., S. Lanning, et al. (2007). The netflix prize. In *Proceedings of KDD cup and workshop*, Volume 2007, pp. 35. New York, NY, USA.

Bourel, M., J. Cugliari, Y. Goude, and J.-M. Poggi (2020, December). Boosting Diversity in Regression Ensembles. working paper or preprint.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Brown, G., J. Wyatt, R. Harris, and X. Yao (2005). Diversity creation methods: a survey and categorisation. *Information Fusion 6*(1), 5–20.

Brown, G., J. L. Wyatt, and P. Tiňo (2005, December). Managing diversity in regression ensembles. *J. Mach. Learn. Res. 6*, 1621–1650.

De Rooij, S., T. Van Erven, P. D. Grünwald, and W. M. Koolen (2014). Follow the leader if you can, hedge if you must. *The Journal of Machine Learning Research 15*(1), 1281–1316.

Debry, E. and V. Mallet (2014). Ensemble forecasting with machine learning algorithms for ozone, nitrogen dioxide and pm10 on the prev'air platform. *Atmospheric environment 91*, 71–84.

Devaine, M., P. Gaillard, Y. Goude, and G. Stoltz (2013). Forecasting electricity consumption by aggregating specialized experts. *Machine Learning 90*(2), 231–260.

Freund, Y. and R. E. Schapire (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences 55*(1), 119–139.

Gaillard, P. and Y. Goude (2014). Forecasting electricity consumption by aggregating experts; how to design a good set of experts. *to appear in Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension*.

Gaillard, P., Y. Goude, and R. Nedellec (2016). Additive models and robust aggregation for gefcom2014 probabilistic electric load and electricity price forecasting. *International Journal of forecasting 32*(3), 1038–1050.

Gaillard, P., G. Stoltz, and T. Van Erven (2014). A second-order bound with excess losses. In *Conference on Learning Theory*, pp. 176–196.

Gourieroux, C. and C. Y. Robert (2006). Stochastic unit root models. *Econometric Theory 22*(6), 1052–1090.

Herbster, M. and M. K. Warmuth (1998). Tracking the best expert. *Machine learning 32*(2), 151–178.

Littlestone, N. and M. K. Warmuth (1994). The weighted majority algorithm. *Information and computation 108*(2), 212–261.

Nowotarski, J. and R. Weron (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews 81*, 1548–1568.

Reeve, H. W. and G. Brown (2018). Diversity and degrees of freedom in regression ensembles. *Neurocomputing 298*, 55 – 68.

Taillardat, M., O. Mestre, M. Zamo, and P. Naveau (2016). Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Monthly Weather Review 144*(6), 2375–2393.

Thorey, J., V. Mallet, and P. Baudin (2017). Online learning with the continuous ranked probability score for ensemble forecasting. *Quarterly Journal of the Royal Meteorological Society 143*(702), 521–529.

Uniejewski, B. and R. Weron (2018). Efficient forecasting of electricity spot prices with expert and lasso models. *Energies 11*(8), 2039.

Vovk, V. (1998). A game of prediction with expert advice. *Journal of Computer and System Sciences 56*(2), 153–173.

Wintenberger, O. (2017). Optimal learning with bernstein online aggregation. *Machine Learning 106*(1), 119–141.

Zamo, M. (2016). *Statistical Post-processing of Deterministic and Ensemble Wind Speed Forecasts on a Grid*. Ph. D. thesis.