

Apprentissage du signal prix de l'électricité. Arbres de régression, séries temporelles et prédictions à long terme

Louis-Victor PASQUIER* Antoine CORNUEJOLS* Suzanne PINSON**

*Agroparistech, UMR 518 Statistiques et Génomes. F-75231 Paris, France

**Université Paris-Dauphine, LAMSADE, CNRS-UMR 7243, PSL 75775 Paris Cedex 16

1 Données et méthode d'apprentissage

Le signal à prédire est le prix spot de l'électricité en €/MWh. Les données se présentent sous la forme d'une série temporelle échantillonnée au pas de temps horaire mesurée sur une année complète, soit un total d'environ 100 000 points relevés. Afin de normaliser les données, le prix d'une heure de consommation est exprimé par rapport au prix moyen de l'année étudiée.

	date	dayofyear	month	dayofweek	hour	we	hol	target
1	2005-01-01 00:00:00	1	1	5	0	1	1	0.3181746
2	2005-01-01 01:00:00	1	1	5	1	1	1	0.3813810
3	2005-01-01 02:00:00	1	1	5	2	1	1	0.3218170

FIG. 1 – Échantillon de l'ensemble d'apprentissage.

Une moyenne pondérée des prédictions de chacun des arbres de régression appris par une méthode de *Random Forests* est employée pour déterminer la variable d'intérêt y .

2 Résultats expérimentaux

La performance des modèles appris est mesurée grâce à la variance expliquée, similaire à un écart quadratique moyen normalisé : $Var.exp. = 1 - \frac{Variance(y_{observée} - y_{modèle})}{Variance(y_{observée})}$. Une variance expliquée de $\approx 0,6$ indique un bon degré de prédictivité.

Ens. app	Ens. test	Nb arbres	Nb attr.	Min. ex./feuille	Var. exp. app.	Var. exp. test
2005-2012	2014	200	5	50	82%	62%
2010-2012	2014	500	5	50	83%	61%
2005-2012	2014	200	5	1	76%	62%

TAB. 1 – Quelques résultats en fonction de divers paramétrages.

Sur une journée Le modèle appris pour une journée (figure 2) prédit bien les heures les moins chères (5h la nuit et 15h en journée) et les heures les plus chères (8h et 18h).