Boosting the rates by adding structural model constraints : application to modeling the electric consomption.

Dominique Picard

Université Paris-Diderot et LPMA

picard@math.univ-paris-diderot.fr

In this talk, we provide a prediction method based on a sparse learning process elaborated on a very high dimensional information, which will be able to include new -potentially high dimensional- influential variables and adapt to different contexts of prediction. We elaborate and test this method in the setting of predicting the national French intra day load curve, over a period of time of 7 years on a large data basis including daily French electrical consumptions as well as many meteorological inputs, calendar statements and functional dictionaries. The prediction box incorporates a huge contextual information coming from the past, organizes it in a manageable way through the construction of a 'smart' encyclopedia of 'scenarios', provides experts elaborating strategies of prediction by comparing the day at hand to referring scenarios extracted from the encyclopedia, and then harmonizes the different experts. The prediction box is built using successive learning procedures and a key component of this construction is an approximation of each day of the past with a large number of explanatory variables using high dimensional methods with sparsity constraints.

Sparsity means that among the potentially explanatory variables, in fact only a small number are truly significant even if we don't know which one. The sparse methods are now quite well understood and working surprisingly well. However, they require some configuration of the explanatory variables. For instance, redundancy in the these variable may cause increasing instability and damage the results. Typical redundancy unfortunately occurs in real data as this is the case for electrical intraday load curves.

An interesting remedy consists in structuring the sparsity, i.e. grouping the explanatory variables into blocks and assuming that only a small number of these blocks are significant.

In some problems, 'natural' blocks can be suggested. If it is not the case, it is interesting to consider different types of grouping and their results in terms of rates of convergence.

We investigate here a 'boosting-blocking-strategy' which aims in a first step

1

at scattering the typically highly correlated variables (leaders) into different blocks and then gather the 'other variables' in the blocks by affinity with the 'leaders'.

This method proves to have good properties in terms of accelerating the rates of convergence. We show that it also gives interesting results in the context of electrical consumption.