Forecasting the electricity consumption by aggregating specialized experts

Pierre Gaillard (EDF R&D, ENS Paris)

with Yannig Goude (EDF R&D)
Gilles Stoltz (CNRS, ENS Paris, HEC Paris)

June 2013 – WIPFOR
Goal

**Short-term** (one-day-ahead) forecasting of the French electricity consumption

Many models developed by EDF R&D: parametric, semi-parametric, and non-parametric

Evolution of the electrical scene in France
⇒ existing models get questionable

Adaptive methods of models aggregation
Setting – Sequential prediction with expert advice

Each instance $t$
- Each expert suggests a prediction $x_{i,t}$ of the consumption $y_t$
- We assign weight to each expert and we predict

$$\hat{y}_t = \hat{p}_t \cdot x_t \left( = \sum_{i=1}^{N} \hat{p}_{i,t} x_{i,t} \right)$$

Our goal is to minimize our cumulative loss

$$\sum_{t=1}^{T} (\hat{y}_t - y_t)^2$$

Our loss

$$\min_{i=1, \ldots, N} \sum_{t=1}^{T} (x_{i,t} - y_t)^2$$

Loss of the best expert

$$R_T$$

Estimation error

Good set of experts

Good aggregating algorithm
Setting – Sequential prediction with expert advice

Each instance $t$
- Each expert suggests a prediction $x_{i,t}$ of the consumption $y_t$
- We assign weight to each expert and we predict

$$\hat{y}_t = \hat{p}_t \cdot x_t \quad \left( = \sum_{i=1}^{N} \hat{p}_{i,t} x_{i,t} \right)$$

Our goal is to minimize our cumulative loss

$$\sum_{t=1}^{T} (\hat{y}_t - y_t)^2 = \min_{q \in \Delta_N} \sum_{t=1}^{T} (q \cdot x_t - y_t)^2 + R_T$$

- Our loss
- Loss of the best convex combination
- Estimation error

Good set of experts
- As varied as possible

Good aggregating algorithm
Minimizing both approximation and estimation error

\[ \sum_{t=1}^{T} (\hat{y}_t - y_t)^2 = \min_{q \in \Delta_N} \sum_{t=1}^{T} (q \cdot x_t - y_t)^2 + R_T \]

**Our loss**

**Approximation error**

**Estimation error**

**Approximation error**

⇒ good heterogeneous set of experts

Ex: specializing the experts, bagging, boosting, . . .

**Estimation error**

⇒ efficient algorithm for aggregating specialized experts

Ex: Exponentially weighted average, Exponentiated Gradient, Ridge, . . .

Prediction Learning and Games, Cesa-Bianchi and Lugosi, 2006
I. Aggregating algorithms

Prediction Learning and Games, Cesa-Bianchi and Lugosi, 2006
Exponentially weighted average forecaster (EWA)

Each instance $t$
- Each expert suggests a prediction $x_{i,t}$ of the consumption $y_t$
- We assign to expert $i$ the weight

$$\hat{p}_{i,t} = \frac{\exp \left( -\eta \sum_{s=1}^{t-1} (x_{i,s} - y_s)^2 \right)}{\sum_{j=1}^{N} \exp \left( -\eta \sum_{s=1}^{t-1} (x_{j,s} - y_s)^2 \right)}$$

- and we predict $\hat{y}_t = \sum_{i=1}^{N} \hat{p}_{i,t} x_{i,t}$

Our cumulated loss is upper bounded by

$$\sum_{t=1}^{T} (\hat{y}_t - y_t)^2 \leq \min_{i=1,\ldots,d} \sum_{t=1}^{T} (x_{i,t} - y_t)^2 + \sqrt{T \log N}$$

- Our loss
- Loss of the best expert
- Estimation error
Exponentially weighted average forecaster (EWA)

Each instance $t$
- Each expert suggests a prediction $x_{i,t}$ of the consumption $y_t$
- We assign to expert $i$ the weight

$$\hat{p}_{i,t} = \frac{\exp(-\eta \sum_{s=1}^{t-1} (x_{i,s} - y_s)^2)}{\sum_{j=1}^{N} \exp(-\eta \sum_{s=1}^{t-1} (x_{j,s} - y_s)^2)}$$

- and we predict $\hat{y}_t = \sum_{i=1}^{N} \hat{p}_{i,t} x_{i,t}$

Our cumulated loss is upper bounded by

$$\sum_{t=1}^{T} (\hat{y}_t - y_t)^2 \leq \min_{q \in \Delta_N} \sum_{t=1}^{T} (q \cdot x_t - y_t)^2 + \text{Estimation error}$$
Motivation of convex combinations

\[ \nabla \psi_t(\hat{p}_t) + \hat{p}_t \]

\[ q^* + \]

\[ \Delta_N \]
Exponentiated gradient forecaster (EG)

Each instance $t$
- Each expert suggests a prediction $x_{i,t}$ of the consumption $y_t$
- We assign to expert $i$ the weight
  \[
  \hat{p}_{i,t} \propto \exp\left(-\eta \sum_{s=1}^{t-1} \ell_{i,s}\right)
  \]
  where $\ell_{i,s} = 2(\hat{y}_s - y_s)x_{i,s}$
- and we predict $\hat{y}_t = \sum_{i=1}^{N} \hat{p}_{i,t}x_{i,t}$

Our cumulated loss is then bounded as follow

\[
\sum_{t=1}^{T} (\hat{y}_t - y_t)^2 \leq \min_{q \in \Delta_N} \sum_{t=1}^{T} (q \cdot x_t - y_t)^2 + \Box \sqrt{T \log N}
\]

Our loss
Loss of the best convex combination
Estimation error

Idea of proof

\[
\sum_{t=1}^{T} (\hat{y}_t - y_t)^2 - (q^* \cdot x_t - y_t)^2 \leq \sum_{t=1}^{T} (\hat{p}_t \cdot x_t - y_t)x_t \cdot (\hat{p}_t - q^*)
\]

\[
= \sum_{t=1}^{T} \ell_t \cdot (\hat{p}_t - q^*)
\leq \sum_{t=1}^{T} \hat{p}_t \cdot \ell_t - \min_i \sum_{t=1}^{T} \ell_{i,t}
\]
II. A good set of experts
Consider as heterogeneous experts as possible

Some ideas to get more variety inside the set of experts

- Consider heterogeneous prediction methods
  - **Gam**: semi-parametric method
  - **KWF**: functional method based on similarity between days
    Clustering functional data using Wavelets, Antoniadis and al, 2013

- Create new experts from the same method thanks to boosting, bagging
- Vary the considered covariate: weather, calendar, ...
- **Specializing the experts**: focus on specific situation (cloudy days,...) during the training
The dataset

The dataset includes 1,696 days from January 1, 2008 to June 15, 2012

- The **electricity consumption** of EDF customers
- **Side information**
  - weather: temperature, nebulosity, wind
  - temporal: date, EJP
  - loss of clients

We remove uncommon days (public holidays ±2) i.e., 55 days each year.

We split the dataset in two subsets

- Jan. 2008 – Aug. 2011: **training set** to build the experts
Performance of the forecasting methods and of the aggregating algorithms

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gam</td>
<td>847</td>
</tr>
<tr>
<td>KWF</td>
<td>1287</td>
</tr>
<tr>
<td>EWA</td>
<td>813</td>
</tr>
<tr>
<td>EG</td>
<td>778</td>
</tr>
</tbody>
</table>
Specializing the experts to diversify

Idea
Focus on specific scenarios during the training of the methods

Meteorological scenarios
- High / low temperature
- High / low variation of the temperature (since the last day, during the day)

Other scenarios
- High / low consumption
- Winter / summer

Such specialized experts suggest prediction only the days corresponding to their scenario
At day $t$, we consider

$$T_t = \text{average temperature of the day}$$

We normalize $T_t$ on $[0, 1]$ and we choose for each day the weight

$$w_t = (1 - T_t)^2$$

We then train our forecasting method using the prior weights $w_t$ on the training days.
Weights given in 2008 for several specializing scenarios

- **Difference of temp. with last day**
- **Hot / cold days**
- **High / low consumption**
- **Variation of temp. during the day**

![Graphs showing the weights for different scenarios](image-url)
Aggregating experts that specialize

Setting

Each day some of the experts are active and output predictions (according to their specialization) while other experts do not

When the expert \( i \) is non active, we do not have access to its prediction

A solution is to assume that non active experts output the same prediction \( \hat{y}_t \) as we do and solve the fixed-point equation

\[
\hat{y}_t = \sum_{j \text{ active}} \hat{p}_{j,t} x_{j,t} + \sum_{i \text{ non active}} \hat{p}_{i,t} \hat{y}_t
\]

Can be extended to activation functions of the experts \( \in [0, 1] \)

Forecasting the electricity consumption by aggregating specialized experts, Devaine and al., 2013
Performance of algorithms with specialized experts

<table>
<thead>
<tr>
<th>Méthode</th>
<th>RMSE (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gam</td>
<td>847</td>
</tr>
<tr>
<td>KWF</td>
<td>1287</td>
</tr>
<tr>
<td>EWA</td>
<td>813</td>
</tr>
<tr>
<td>EG</td>
<td>778</td>
</tr>
</tbody>
</table>

| Spec + EWA | 765       |
| Spec + EG  | 714       |

**Diagram:**

- **Gam**
- **KWF**
- **EWA**
- **EG**
- **Specified Gam**
- **Specified KWF**

Weights range from 0.00 to 0.25.
Performance of algorithms with specialized experts

### Hour

- **Gam**
- **KWF**

<table>
<thead>
<tr>
<th>Hour</th>
<th>RMSE (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>500</td>
</tr>
<tr>
<td>8</td>
<td>1000</td>
</tr>
<tr>
<td>12</td>
<td>1500</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>28</td>
<td>16</td>
</tr>
<tr>
<td>32</td>
<td>20</td>
</tr>
</tbody>
</table>

### Month

- **EWA**
- **EG**
- **Spec.EWA**
- **Spec.EG**

<table>
<thead>
<tr>
<th>Month</th>
<th>RMSE (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>500</td>
</tr>
<tr>
<td>Feb</td>
<td>1000</td>
</tr>
<tr>
<td>Mar</td>
<td>1500</td>
</tr>
<tr>
<td>Avr</td>
<td>4</td>
</tr>
<tr>
<td>May</td>
<td>8</td>
</tr>
<tr>
<td>Jun</td>
<td>12</td>
</tr>
<tr>
<td>Sep</td>
<td>16</td>
</tr>
<tr>
<td>Oct</td>
<td>20</td>
</tr>
<tr>
<td>Nov</td>
<td>500</td>
</tr>
<tr>
<td>Dec</td>
<td>1000</td>
</tr>
</tbody>
</table>