The Experts Method for the prediction of periodic multivariate time series of high dimension

Il Metodo degli Esperti per la previsione di serie temporali multivariate e periodiche, di dimensione elevata

Giacomo Aletti, Marco Bellan and Alessandra Micheletti

Abstract We propose a method, called *Experts Method*, to predict the evolution of a high dimensional multivariate set of time series. The method is based on the definition of a set of "experts", which are portions of a training set of the considered time series which best fit the data immediately preceding those to be predicted. A suitable combination of Singular Value Decompositions is used to filter out the noise, and provide robust predictions. The advantage of this method, if compared with classical multivariate time series analysis, is that it can be applied also when the time series column order is reshuffled, from time to time, in the collected dataset.
Abstract *In questo lavoro viene proposto un metodo, detto* Metodo degli Esperti*, per predire l'evoluzione di un insieme multivariato e di grosse dimensioni di serie temporali. Il metodo é basato sulla definizione di un insieme di "esperti", ovvero di porzioni, di un training set delle serie temporali considerate, che approssimano al meglio i dati che precedono quelli che devono essere previsti. Viene utilizzata una opportuna combinazione di decomposizioni ai valori singolari per filtrare il rumore, e fornire previsioni robuste. Il vantaggio di questo metodo, rispetto ai classici metodi di analisi di serie temporali multivariate, é che esso puó essere applicato anche quando l'ordine con cui le serie sono registrate nelle colonne del dataset, viene scambiato di tanto in tanto.*

Key words: SVD, prediction, multivariate time series

Giacomo Aletti, Alessandra Micheletti
Dept. of Environmental Science and Policy, Università degli studi di Milano e-mail: `giacomo.aletti@unimi.it`, `alessandra.micheletti@unimi.it`

Marco Bellan
Dept. of Mathematics, Università degli studi di Milano e-mail: `marco.bellan@studenti.unimi.it`

1 Introduction

In this paper we describe a method, called *Experts Method*, to forecast the evolution of a multivariate set of time series, of big dimension, and with partially censored data. It has been applied to the data provided by the H2020 Big Data Horizon Prize 2017 [1]. The data subject to the forecast are energy flow-related measurements over $N_T = 1912$ high-tension lines registered over two years (called *target data*). The lines have been anonymized with respect to location, thus they are provided as purely temporal data. The energy flows are registered every 5 minutes for one year, but since the data recording can be switched off for some (random) period of time, some of the time series show missing data. The data are divided into subsequent files, where each column of the files contains the observations of one high-tension line in the time period $[t, t + \Delta t]$. The length Δt of the periods of observation can be different from file to file (it anyway is always bigger or equal than 1 hour). This setting emulates the fact that in real applications data are recorded continuously by sensors, but they can be passed to the statistical analysis during different times, either regularly or when the measurements overcome a fixed threshold, or some extra control is planned and an immediate forecast of the future behaviour is needed.

The Experts Method here proposed is thus suitable to predict the evolution of a multivariate set of *data streams* of big dimension, where the observed streams are given during different random times.

The method here proposed is supervised, thus the target data have been divided into a training set, to set up the method, and an adapt set, on which the method is tested. Anyway, differently from deep learning methods based on neural networks, our method does not need a huge training set to be trained. Moreover our method allows to predict the behaviour of the time series even if their order in the columns of the provided files is reshuffled from time to time.

The presence of missing data in the target has been taken into account as follows:

- in the training set the NaN are substituted with zeros, the +inf values (if any) are substituted with the maximum floating point value in double precision, the -inf values (if any) are substituted with the minimum floating point value in double precision.
- In the adapt set if the NaN's sequence is in between two observed values of the column in the file, we interpolate linearly the missing data between the two observed values. If the NaN's sequence is starting at the beginning of the file, we fill all the missing values with the first observed value in the column. If the NaN sequence is located at the end of the file, we fill the missing values with the last observed value.

In the following we describe the theoretical foundation of the Expert Method, and the implementation scheme.

2 Weigthed Sobolev discrete space

Given two bounded, piece-wise differentiable functions f and g on a fixed interval $[t_0, t_1]$, we will measure the distance between them through the weighted Sobolev distance d_S so defined:

$$d_S^2(f,g) := \int_{t_0}^{t_1} (f(t)-g(t))^2\,dt + w\int_{t_0}^{t_1} (f'(t)-g'(t))^2\,dt. \tag{1}$$

We will often refer to sets of equally sampled functions $\{f_j(t_i), 0 \leq i < M, 0 \leq j < N\}$, which form matrices $F = [f_j(t_i)]_{i,j}$. Thus, in our application, M is the number of considered time steps, N is the number of observed high-tension lines and $f_j(t_i)$ is the energy flow measured on line j at time t_i. The matrix $\Delta F = [f_j(t_i) - f_j(t_{i-1})]_{i,j}$ will play the role of estimating the piecewise constant derivative of each function and will be appended to F to form the extended matrix $F^{(\text{ext})}$:

$$F^{(\text{ext})} = \begin{pmatrix} F \\ \alpha\Delta F \end{pmatrix}, \qquad \text{where } \alpha^2 = w(1-\tfrac{1}{M}).$$

With this notation, given two matrices $F^{(1)}$ and $F^{(2)}$ of the same form, the Froebenious norm of the difference between the two extended matrices relates to the sum of the Sobolev squared distance of the corresponding functions:

$$\begin{aligned}
&\|F^{(1)\,(\text{ext})} - F^{(2)\,(\text{ext})}\| \\
&\qquad = \sum_j \Big(\sum_{i<M} (f_j^{(1)}(t_i) - f_j^{(2)}(t_i) \Big)^2 \\
&\qquad\qquad + \alpha^2 \sum_{0<i<M} \big((f_j^{(1)}(t_i) - f_j^{(1)}(t_{i-1})) - (f_j^{(2)}(t_i) - f_j^{(2)}(t_{i-1})) \big)^2 \\
&\qquad \simeq \sum_j \Big(\int_{t_0}^{t_1} (f_j^{(1)}(t) - f_j^{(2)}(t))^2\,dt + w\int_{t_0}^{t_1} ({f^{(1)}}'_j(t) - {f^{(2)}}'_j(t))^2\,dt \Big) \\
&\qquad = \sum_j d_S^2(f_j^{(1)}, f_j^{(2)}).
\end{aligned}$$

We will use the Sobolev squared distance to identify the subset of the training set which fits better with the last observations of the electric lines before the prediction.

3 Implementation of the Experts Method

We assume to receive in input a matrix B of dimension $M_A \times N_A$

$$B = \{B_{t_i,j}\}, \qquad 0 \le i < M_A, 0 \le j < N_A,$$

where again M_A is the number of recorded time steps, and N_A is the (total) number of electric lines. We want to predict $\{B_{t_i,j}, M_A \le i < M_A + \text{Hor}, 0 \le j < N_A\}$, that is we want to predict the lines energy flow in a number Hor of time steps forward.

We extract the days of the week, the hours and the minutes of time spanned by B, and we extract from the training set the family of all the observations, called "*experts*" , $E = \{E^{(k)}\}$ that insist on the same timespan, but in another year and month,

$$E^{(k)} = \{E^{(k)}_{t_i,j}\}, \qquad 0 \le i < M_A + \text{Hor}, 0 \le j^* < N_T, 0 \le k < N_{\text{exp}}.$$

In this way we form a sample of "experts", based on the observation that our data show a periodicity, with a period of one week. This is reasonable, since the variation in the request of electric energy is mainly related with the working habits of a region and their weekly recurrence. We neglect here the effects of yearly seasonality (differences from summer and winter), because we have a training set composed only by one year of observations. Also this type of seasonality could be taken into account in presence of richer datasets.

We do not assume that $N_T = N_A$, since the recording of some lines can have been switched off and we also assume that the columns of B can be reshuffled in subsequent files. The Experts Method guarantees a good prediction performance also when the recorded lines are stored into different columns of the data matrix at different time frames.

3.1 Approximation by experts

Given an expert $E^{(k)}$, we use

1. the first part $E^{(k,1)} = \{E^{(k)}_{t_i,j}, 0 \le i < M_A, 0 \le j < N_T\}$ to fit $\{B_{t_i,j}, 0 \le i < M_A, 0 \le j < N_T\}$.
2. the second part $E^{(k,2)} = \{E^{(k)}_{t_i,j}, M_A \le i < M_A + \text{Hor}, 0 \le j < N_T\}$ to predict the unknown $\{B_{t_i,j}, M_A \le i < M_A + \text{Hor}, 0 \le j < N_T\}$.

We compute the extended matrices of experts $E^{(k,1)}{}^{(\text{ext})}$, and then, in order to fit each line $\{B_{t_i,j}, 0 \le i < M_A\}$ (without overfitting it), we take the best rank $r < \min(N_T, M_A)$ approximation of each extended expert by computing the SVD decomposition [2, 3] of each $E^{(k,1)}{}^{(\text{ext})}$. Accordingly, given

$$E^{(k,1)}{}^{(\text{ext})} = \sum_{v=1}^{\min(N_T,M_A)} \mathbf{u}_v^{(k)} s_v^{(k)} \mathbf{v}_v^{(k)}{}^{\top}$$

with $s_1^{(k)} \geq s_2^{(k)} \geq \cdots \geq s_{\min(N_T,M_A)}^{(k)} \geq 0$, we first remove the unnecessary noise (where $\frac{s_v^{(k)}}{s_1^{(k)}} < factor$). Then we keep only the first r components of the SVD decomposition, defining the smoothed experts

$$S_E^{(k,1)\,(\mathrm{ext})} = \sum_{v=1}^{r} \mathbf{u}_v^{(k)} s_v^{(k)} \mathbf{v}_v^{(k)\,\top}.$$

We note that $S_E^{(k,1)}$ is the best rank r approximation, using the Sobolev distance, of the functions on the columns of each $E^{(k,1)}$.

3.2 Prediction from each expert

We predict each column of B independently of the others, thus we fix the column $\tilde{\mathbf{B}}_j = \{B_{t_i,j}, 0 \leq i < M_A\}$ and we compute the extended column $\tilde{\mathbf{B}}_j^{(\mathrm{ext})} = \{B_{t_i,j}^{(\mathrm{ext})}, 0 \leq i < M_A\}$. Each smoothed expert is requested to fit linearly $\tilde{\mathbf{B}}_j$ with its best $\mathtt{RED}$-rank linear combination, where $\mathtt{RED} \leq r$ is a further parameter chosen to reduce the dimensionality of the problem. Accordingly, given $X_{v,j}^{(k)} = \mathbf{u}_v^{(k)} \cdot \tilde{\mathbf{B}}_j^{(\mathrm{ext})}$, each $X_{v,j}^{(k)}$ represents the projection of $\tilde{\mathbf{B}}_j^{(\mathrm{ext})}$ on the ortonormalized vectors $\mathbf{u}_v^{(k)}$. Then, for any k, j, we take the $\mathtt{RED}$ indices $\{v_{j,1}^{(k)}, \ldots, v_{j,\mathtt{RED}}^{(k)}\} \subseteq \{1, \ldots, r\}$ which maximize $(X_{v,j}^{(k)})^2$, and we solve the least squares problem

$$\mathbf{x}_j^{(k)} = \operatorname*{arg\,min}_{\mathbf{x} \in \mathbb{R}^{N_T}} \Bigg\| \underbrace{\left(\sum_{l=1}^{\mathtt{RED}} \mathbf{u}_{v_{j,l}^{(k)}}^{(k)} s_{v_{j,l}^{(k)}}^{(k)} \mathbf{v}_{v_{j,l}^{(k)}}^{(k)\,\top} \right)}_{\mathtt{RED}\text{ approx. of } S_E^{(k,1)\,(\mathrm{ext})}} \cdot \mathbf{x} - \tilde{\mathbf{B}}_j^{(\mathrm{ext})} \Bigg\|^2.$$

The solution has the analytic expression $\mathbf{x}_j^{(k)} = \sum_{l=1}^{\mathtt{RED}} \frac{X_{v_{j,l}^{(k)},j}^{(k)}}{s_{v_{j,l}^{(k)}}^{(k)}} \mathbf{v}_{v_{j,l}^{(k)}}^{(k)}$ for each expert k. A first prediction of $\{B_{t_i,j}, M_A \leq i < M_A + \mathrm{Hor}\}$ is hence based on $E^{(k,2)}$ and it is given by

$$\check{B}_{t_i,j}^{(k)} = \sum_{j^*} E_{t_i,j^*}^{(k)} x_{j^*,j}^{(k)}, \qquad M_A \leq i \leq M_A + \mathrm{Hor}.$$

Finally, to ensure that the prediction will start from the last observed value in the adapt file, a correction term is added $\hat{B}_{t_i,j}^{(k)} = \check{B}_{t_i,j}^{(k)} + (B_{t_{M_A}-1,j}^{(k)} - \sum_{j^*} E_{t_{M_A}-1,j^*}^{(k)} x_{j^*,j}^{(k)}) * h(t_i - t_{M_A} - 1)$, where h is a decreasing non-negative function with $h(0) = 1$.

3.3 Final prediction

For each column j of B, each expert k has given its prediction $\{\hat{B}^{(k)}_{t_i,j}, M_A \leq i < M_A + \text{Hor}\}$. Among them, we extract the best nEXP experts (with $\texttt{nEXP} \leq N_{\text{exp}}$) according to their strength in the linear fitting of $\tilde{\mathbf{B}}_j$. In other words, we extract the indices $\{k_{j_1}, \ldots, k_{j_{\texttt{nEXP}}}\}$ that maximize $\left\{\sum_{l=1}^{\texttt{RED}} \big(X^{(k)}_{v^{(k)}_{j,l},j}\big)^2, k = 1, \ldots, N_{\text{exp}}\right\}$. The final prediction is given robustly as the median of the predictions of these experts $\hat{B}_{t_i,j} = \text{median}\left\{\hat{B}^{(k)}_{t_i,j}, k \in \{k_{j_1}, \ldots, k_{j_{\texttt{nEXP}}}\}\right\}$. The Experts method here proposed is thus based on the set of parameters α,r (or, equivalently, $factor$), RED, nEXP, which must be tuned to balance the computational costs and the precision of the prediction.

4 Numerical results

The method has been applied to predict iteratively 12 time steps forward, for each input matrix B of the test set. The method was not much sensitive to α, thus we fixed $\alpha = 1$, and nEXP has been fixed to 11, since it was the maximum value which guaranteed us to remain in the maximum allowed execution time in the H2020 competition. We made the parameters $factor$ and RED vary, and we computed the corresponding MSE, and the ratio MSE(expert method)/MSE(constant prediction). The results are reported in the tables below.

MSE

RED	3	5	7	9	12
factor					
0.00001	935	900	875	876	944
0.00010	935	900	875	876	944
0.00100	933	901	874	875	944
0.01000	926	929	930	952	938
0.10000	1250	1250	1250	1250	1250

RATIO

RED	3	5	7	9	12
factor					
0.00001	0.924827	0.890208	0.865480	0.866469	0.933729
0.00010	0.924827	0.890208	0.865480	0.866469	0.933729
0.00100	0.922849	0.891197	0.864491	0.865480	0.933729
0.01000	0.915925	0.918892	0.919881	0.941642	0.927794
0.10000	1.236400	1.236400	1.236400	1.236400	1.236400

References

1. H2020 inducement prize: Big Data technologies 2017. http://ec.europa.eu/research/horizonprize/index.cfm?prize=bigdata
2. Sudipto, B., Anindya, R.: Linear Algebra and Matrix Analysis for Statistics, CRC Press, 2014.
3. Yanai, H., Takeuchi, K., Takane, Y.: Projection matrices, generalized inverse matrices, and singular value decomposition, Springer, 2011.