

On the use of the wavelet decomposition for time series prediction

Alan Balendran
Séries Temporelles
SORBONNE UNIVERSITÉ

April 20, 2020

1 Introduction

En machine learning, l'étude des séries temporelles se différencie par le fait que les observations ne sont pas indépendantes. En effet les observations sont indexées selon une certaine temporalité. De ce fait certaines techniques de machine learning doivent être adapté, comme par exemple la technique de *cross validation* qui permet de choisir un paramètre en découpant le jeu de d'observation aléatoirement en plusieurs sous-ensembles, dans l'étude de série temporelles, cette découpe doit être structuré en préservant la temporalité. Un autre problème est celui de sous/sur apprentissage, dans le premier cas notre modèle ne réussira pas à apprendre de nos données et ne pourra donc pas présenter de bonne performances lors de prédictions tandis qu'un sur-apprentissage aboutira à un modèle qui se sera totalement adapté aux données (cela inclut notamment le bruit) mais ne pourra pas s'adapter à de nouvelles données.

L'article traite justement d'une méthode qui présente un bon compromis.

2 Présentation

L'article traite d'une méthode de prédiction sur des données décomposé à l'aide d'ondelettes a priori et en sommant ensuite les différentes prédictions. De plus, l'usage de réseaux de neurones pour la prédictions est expliqué par le fait qu'on suppose que la relation entre \hat{x}_{k+p}^* (l'observation à prédire) et les observations antérieures x_1, x_2, \dots, x_ℓ (observations observées) est non-linéaire.

Comme nous le savons, dans une série temporelles, étant donnée les x_1, x_2, \dots, x_ℓ l'observation x_{k+p} est obtenu à partir des observations antérieurs à celle ci. Plus précisément elle est obtenu en minimisant le risque suivant:

$$C_{\text{gen}} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^T \mathbb{E} \{ (\hat{x}_{k+p} - x_{k+p})^2 | x_k, x_{k-1}, \dots \}$$

qui est minimal pour

$$\hat{x}_{k+p}^* = \mathbb{E} \{ x_{k+p} | x_k, x_{k-1}, \dots \}$$

. Nos observations étant fixés nous minimiserons plutôt la quantité suivante qui est le risque empirique:

$$C_{\text{cmp}} = \frac{1}{l} \sum_{i=1}^l (x_i - \hat{x}_i)^2$$

Une première approche, classique serait d'apprendre directement sur le jeu de données. Un des problèmes de cette méthode est celui de sur/sous apprentissage, le sur apprentissage s'explique par la présence de bruit dans le signal, le modèle va apprendre le bruit et donc faire de mauvaise prédiction, de la même façon un modèle qui n'apprend pas les variations rapides ne va pouvoir prédire de manière assez précise les oscillations.

C'est là où intervient la décomposition en ondelettes.

La décomposition en ondelettes d'un signal permet l'analyse de ce signal à différentes fréquences, permettant ainsi séparer le signal sur plusieurs niveaux, un niveau à basse oscillation qui est exempté de bruit et des niveaux à haute fréquence pour des signaux avec plus de détails.

La construction de filtre pour cette décomposition impose des contraintes afin de permettre la reconstitution du signal initial.

La base de Haar est un candidat respectant les contraintes de reconstitution.

Après avoir décomposé le signal sur différents niveaux, vient l'étape du choix de modèle. Un estimateur est ensuite construit pour chaque niveau.

Remark

- Une méthode est de choisir à chaque niveau un estimateur basée seulement sur le signal de ce niveau, c'est à dire :

$$\begin{aligned}\hat{c}_{Nk+p} &= f_0(c_{N,k}, c_{N,k-1}, \dots, c_{Nk-r_0}) \\ \hat{d}_{j,k+p} &= \hat{f}_j(d_{j,k}, d_{j,k-1}, \dots, d_{j,k-r_j}), \quad j = 1, \dots, N\end{aligned}$$

- Une autre méthode proposé est d'utiliser la dépendance entre chaque série, donnant ainsi les estimateurs suivants :

$$\begin{aligned}\hat{c}_{Nk+p} &= \hat{f}_0(c_{N,k}, \dots, c_{N,k-r}, \dots, d_{j,k}, \dots, d_{j,k-r}, \dots) \\ \hat{d}_{j,k+p} &= \hat{f}_j(d_{j,k}, \dots, d_{j,k-r}, \dots, c_{N,k}, \dots, c_{Nk-r}), \quad j = 1, \dots, N\end{aligned}$$

Cependant cela augmente la dimension des paramètres.

Au final, la prédiction obtenu correspond à la somme des différentes prédictions obtenu à chaque niveau. (grâce aux contraintes de reconstitution imposées).

La prédiction obtenu par cette méthode présente de meilleur résultat que celle par l'approche classique en terme de risque empirique.

Ce résultat se montre de la façon suivante:

Soit f et f_0, \dots, f_j les modèles obtenus respectivement par la méthode classique et celles par décomposition en ondelettes :

$$\hat{f} = \sum_{i=1}^s w_i \varphi \left(\sum_{l=0}^{r-1} a_{i,l} x_{r-l} + a_{i,r} \right)$$

avec $w_i, i = 1, \dots, s, a_{j,l}, j = 1, \dots, s, l = 0, \dots, r$ les coefficients du modèle.

De même nous avons:

$$\hat{f}_0 = \sum_{i=1}^s w_i^0 \varphi \left(\sum_{l=0}^{r-1} a_{0,l,i}^0 c_{N,r-l} + \sum_{n=1}^N \sum_{l=0}^{r-1} a_{n,l,i}^0 d_{n,r-l} + a_{i,r}^0 \right)$$

...

$$\hat{f}_j = \sum_{i=1}^s w_i^j \varphi \left(\sum_{l=0}^{r-1} a_{0,l,i}^j c_{N,r-l} + \sum_{n=1}^N \sum_{l=0}^{r-1} a_{n,l,i}^j d_{n,r-l} + a_{i,r}^j \right)$$

avec $w_i^j, i = 1, \dots, s, j = 0, \dots, N, a_{n,l,i}^j, j = 0, \dots, N, l = 0, \dots, r, n = 1, \dots, N$ les coefficients du modèle.

On remarquera que pour

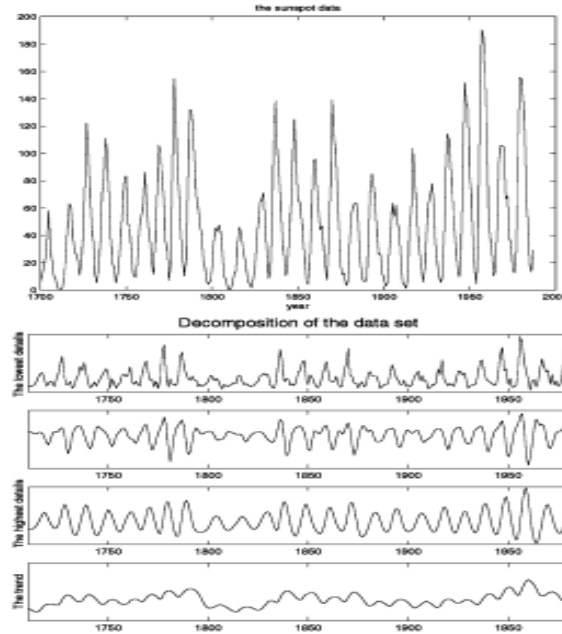
$$a_{n,l,i}^0 = a_{n,l,i}^j, \quad n = 0, \dots, N, \quad l = 0, \dots, r-1, \quad i = 1, \dots, s$$

\hat{f} est une combinaison linéaire des $\hat{f}_0, \dots, \hat{f}_j$. Donc l'ensemble des estimateurs de l'espace C_{emp} est un sous ensemble de C_{emp}^w , ainsi on en déduit que :

$$\min C_{emp}^w \leq \min C_{emp}$$

3 Application

: Le jeu de données utilisé est ici une séries représentant le nombre de tache solaire annuel de 1700 à 1979.



On peut voir qu'en décomposant le signal à différentes échelles, on arrive à distinguer les différentes tendances. Il est alors possible d'apprendre à partir de chaque signal obtenu à l'aide d'un réseau de neurones.

Les différents modèles testés sont:

- **TAR** (Threshold AutoRegressive), un modèle non-linéaire utilisé dans l'étude des séries temporelles.
- Réseau de neurones

- **TTDN** (Time Delay Neural Network), un réseau de neurones utilisé en traitement du langage et qui à la différence d'un réseau de neurones classique, utilise l'information passée des couches de neurones précédentes.
- la méthode présentée.

Table 1
Normalized error obtained on the test sets by using different methods. First column contains the results on the test set 1921–1954 while the second contains those on the test set 1955–1979

Model	$e_{1921-1954}$	$e_{1955-1979}$
TAR	0.097	0.28
Neural nets	0.086	0.35
TDNN	0.093	0.246
Our method	0.076	0.23

On peut voir que la méthode de décomposition présente de meilleur résultat qu'un réseau de neurones et d'autres méthodes sur les données 'brutes'. Ici les performances sont calculés selon l'erreur normalisé sur deux jeux de données test.

4 Conclusion

L'article présente donc une méthode de prédiction basé sur une décomposition à l'aide d'ondelettes, permettant de produire de meilleures prédictions. En effet cette méthode répond à la problématique du sous et sur-apprentissage qui est un problème souvent rencontré en Machine Learning.