

Defi IA

INSA Toulouse
Department de Mathématiques Appliquées

December 2021

Table of Contents

1 Introduction

2 Traitement de données

3 Choix du réseau

4 Conclusion

Introduction

Objectif du défi: Prédire aux stations d'observations les cumuls journaliers de précipitation

Données disponibles sur Kaggle:

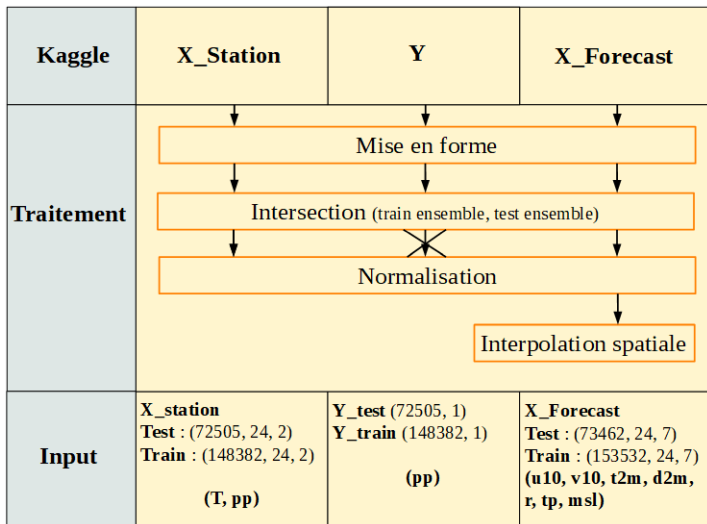
- ▷ Observations aux stations
- ▷ Prévisions des modèles
- ▷ Baseline (utilisée pour évaluer nos réseaux)

Table of Contents

- 1 Introduction
- 2 Traitement de données**
- 3 Choix du réseau
- 4 Conclusion

Traitement de données

Traitement de données

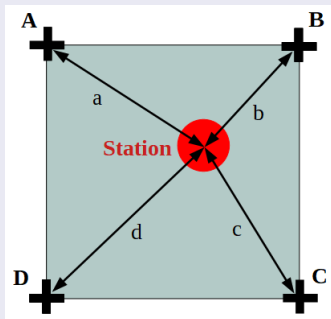


X_forecast

A partir des coordonnées de chaque station de mesure on en a déduit les coordonnées des mailles correspondantes ainsi que les points de grille voisins à chaque station.

Interpolation spatiale

Les distances entre le point de la station et les points de grille voisins sont calculées.



On a ainsi pour la prevision à la station:

$$dist = \frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2} + \frac{1}{d^2}$$

$$prev = \left(\frac{prevA}{a^2} + \frac{prevB}{b^2} + \frac{prevC}{c^2} + \frac{prevD}{d^2} \right) / dist$$

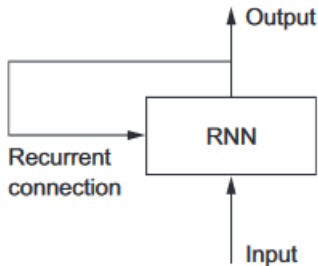
Table of Contents

- 1 Introduction
- 2 Traitement de données
- 3 Choix du réseau**
- 4 Conclusion

Choix du réseau

Les RNN (Recurrent neural network)

Pour ce défi, on a choisit d'utiliser des RNN.



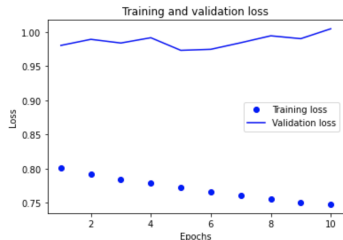
Un RNN traite les séquences en itérant sur les éléments de séquence et en maintenant un état contenant des informations relatives à ce qu'il a vu jusqu'à présent.
→ Un RNN possède une boucle interne.

On a voulu traiter le problème à l'aide de réseaux de neurones.

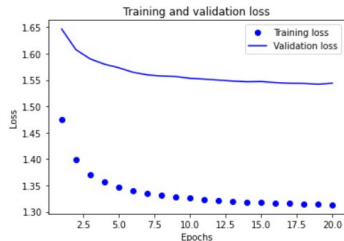
Différents réseaux testés

- Réseau de neurones Dense, inputs : $X_{station}$
- Réseau de neurones LSTM, inputs : $X_{station}$, $X_{forecast}$
- Réseau de neurones LSTM, inputs : $X_{station}$
- Réseau de neurones LSTM, inputs : $X_{forecast}$

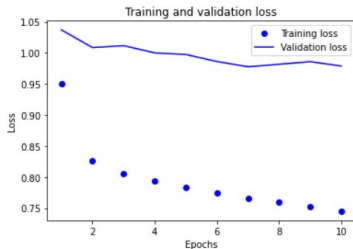
Entraînement des quatre réseaux



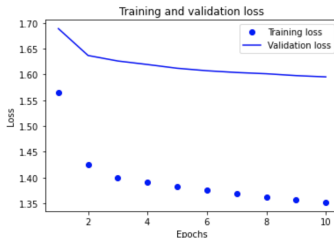
(a) NN Dense , Xstation



(b) NN LSTM , Xstation Xforecast



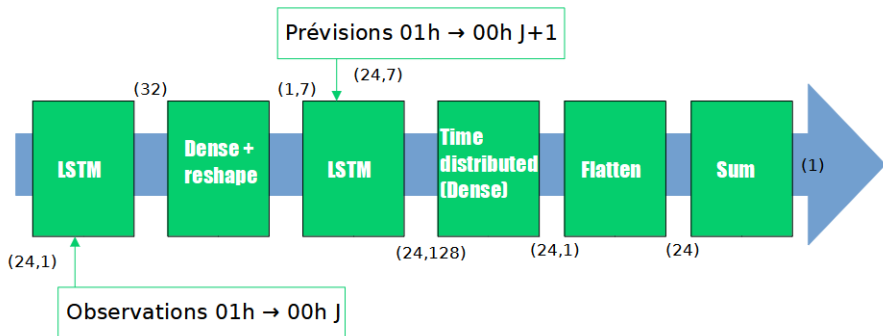
(c) NN LSTM , Xstation



(d) NN LSTM , Xforecast



Architecture du réseau



Optimisation du paramètre du réseau

Paramètres	Valeurs		
Loss	mae	mse	logcosh
Batch size	100	200	400
Epoch numbers	5	10	20

Mean Absolute Percentage Error (MAPE)

La mesure d'évaluation pour ce concours est le pourcentage d'erreur absolu moyen. Le score MAPE est étroitement lié à un score souvent utilisé dans Météo France pour évaluer la prédiction. Il est donné par :

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

où chaque prévision (F_t) est comparée à la valeur réelle (A_t) pour le t -ième élément parmi n éléments.

Table of Contents

- 1 Introduction
- 2 Traitement de données
- 3 Choix du réseau
- 4 Conclusion**

Conclusion

Conclusion

Score des réseaux

Réseau	Score (en interne)
<u>RNN</u> Dense: X_station	44%
<u>RNN</u> LSTM: X_station	31%
<u>RNN</u> LSTM: X_station et X_forecast	67%
<u>RNN</u> LSTM: X_forecast	74%

- Le meilleur score obtenu n'est pas celui du réseau initialement choisi;
- Difference de score sur Kaggle et en interne: échantillon d'évaluation différent.

Merci pour votre attention!