

Atelier permanent de l'OMP et COMET SIL et TSI du CNES

Introduction et panorama du Machine Learning (apprentissage automatique) pour les sciences de l'univers et de la terre.

Programme

09h00 à 12h30 : Mathieu Fauvel/CESBIO/OMP et Florent Chatelain/GIPSA-LAB -présentation des concepts et méthodes les plus couramment utilisées en machine learning. Focus sur la mise en œuvre pratique de ces méthodes dans le domaine des sciences de l'univers de la terre.

12h30 -14h00 Buffet

14h00 - 14h40 : Mauro Dalla Mura/GIPSA-LAB/INP-Grenoble : "Automatic classification of natural signals for environmental monitoring". 14h40 - 15h20 : Syvie Thiria/LOCEAN/IPSL : Apport de l'apprentissage non supervisé à la modélisation en Sciences de l'environnement.

15h20-15h50 : Pause

15h50 - 16h30 : Sylvain Fuertes et Stéphane D'Escrivan/CNES : maintenance prédictive des équipements du satellite (opérations satellites)

16h30 - 17h10 : Olivier Mestre/DIROP/METEO-France : Débiaisage des modèles de prévision numérique du temps par des techniques d'apprentissage automatique, applications à la prévision de la température et de la nébulosité

17h10 - 17h30 : Rémy cabanac/IRAP/OMP : Applications de PCA et SVN à la séparation étoile-galaxie.

COMITE DE PROGRAMME

Hervé Carfantan IRAP/OMP pour l'atelier permanent de l'OMP.

Etienne Gondet UMS831/OMP pour l'atelier permanent de l'OMP.

Nemesio Rodriguez-Fernandez CESBIO/OMP pour l'atelier permanent de l'OMP.

Guillaume Eynard-Bontemps:CNES pour le COMET SIL.

Vincent Martin/CNES pour COMET TSI.

Yannick Tanguy/DSO/SI/2A/CNES pour le COMET TSI.

Programme détaillé

MATINEE METHODOLOGIQUE

- Matthieu Fauvel/CESBIO/OMP et Florent Chatelain (du GIPSA-Lab) .
 - Titre : Présentation des concepts et méthodes les plus couramment utilisées en machine learning. Focus sur la mise en œuvre pratique de ces méthodes dans le domaine des sciences de l'univers de la terre.
 - Résumé : Durant cette matinée, nous présenterons les concepts de base de l'analyse de données et de l'extraction automatique d'information. Les notions d'apprentissage supervisé et non-supervisé, de modèle paramétrique et non-paramétrique ainsi que le sur/sous-apprentissage et la validation de modèle seront introduites. Nous nous focaliserons sur la tâche de classification de données (pixels, signaux, etc ...). Plusieurs méthodes seront ensuite présentées et analysées sur des cas simples. Les méthodes supervisées considérées seront: classifieur Bayésien Naïf (paramétrique ou non), l'analyse discriminante linéaire et quadratique, les k- plus proches voisins, les SVM et les forêts aléatoires. Les méthodes non supervisées, ou dites de clustering, seront: le K-means et les modèles de mélanges Gaussien.
 - Les documents de présentation ainsi que les codes python nécessaires aux simulations seront fournis aux participants.

APRES MIDI : RETOURS D'EXPERIENCES ET METHODES POUR LA RECHERCHE.

- Mauro Dalla-Mura :
 - Titre : Automatic classification of natural signals for environmental monitoring
 - Résumé : This presentation will give an overview of a practical application of machine learning for the automatic classification of environmental signals such as volcano seismic and underwater acoustics acquisitions.
- Sylvie Thiria LOCEAN
 - Titre : Apport de l'apprentissage non supervisé à la modélisation en Sciences de l'environnement.
 - Résumé : L'exposé montrera à l'aide d'exemples en télédétection et en climatologie la portée des algorithmes neuronaux et des modélisations statistiques.
- Sylvain Fuertes et Stéphane D'Escrivan : maintenance prédictive des équipements du satellite (opérations satellites)
- Olivier Mestre : MétéoFrance : Direction des Opérations pour la Prévision
 - Titre : Débiaisage des modèles de prévision numérique du temps par des techniques d'apprentissage automatique, applications à la prévision de la température et de la nébulosité
 - Résumé : La prévision en météorologie se fait classiquement par de gros modèles de Prévision Numérique du Temps (PNT). Ce sont des modèles physiques, basés sur la résolution de l'équation de Navier-Stokes. Ces modèles représentent correctement la dynamique de l'atmosphère, mais sont généralement biaisés localement. Par ailleurs, des processus thermodynamiques complexes comme la formation de nuages sont plus difficiles à modéliser. Afin de corriger localement ces prévisions, on utilise des archives d'observations, et l'on construit des modèles de régression entre données observées et prédicteurs issus de la PNT. C'est un problème classique d'apprentissage supervisé, l'hypothèse étant que les prévisions de la PNT recèlent de l'information sur leur propre erreur. On montrera l'efficacité de diverses techniques d'apprentissage comme les forêts aléatoires, le boosting, et les contraintes qu'elles imposent lorsqu'on les utilise pour faire de la prévision sur de gros volumes de données. On abordera également les techniques de calibration des systèmes de prévision d'ensemble permettant de faire des prévisions probabilistes. Quelques cas d'études seront présentés, comme les prévisions de température et les prévisions de nébulosité pour les opérateurs satellites.
- Rémy Cabanac

- Titre: Applications de PCA et SVN à la séparation étoile-galaxie.
- Résumé: Pouvoir séparer les objets résolus des objets non-résolus sur les images du ciel profond est important pour la mission spatiale Euclid. Je présenterai les algorithmes de séparation testés parmi lesquels, une méthode paramétrique, une méthode PCA et une méthode SVN.