



Rencontres R & D 2019

L'INTELLIGENCE ARTIFICIELLE

La théorie

Panorama historisé de l'Apprentissage Statistique

Philippe Besse - INSA-Toulouse

Big data, data science, machine learning, deep learning, intelligence artificielle... un buzzword ou un battage médiatique en chasse un autre. L'exposé vise à montrer comment la convergence entre grandes masses de données, moyens de calcul et "nouveaux" algorithmes d'apprentissage conduit, certes à des résultats spectaculaires, mais des résultats qu'il est bon de relativiser. Après avoir rapidement décrit l'évolution des grandes familles d'algorithmes d'apprentissage statistique, nous concluons sur une approche nécessairement hybride de la modélisation entre physique ou probabilités et apprentissage issu des données.

Etude d'un exemple concret d'apprentissage profond

Mathieu Serrurier - UPS/IRIT

Dans cet exposé, nous étudierons un cas concret de deep learning sur des images en allant du formatage des données à l'évaluation du modèle obtenu en passant par le paramétrage du modèle et de l'algorithme d'apprentissage.

Quelques projets

Le LAB IA, le FATP et les moyens de calcul

Lior Perez - DSI/DEV

Grâce au Fonds de Transformation de l'Action Publique (FTAP), Météo-France crée un Lab IA. Démarré en février 2019 et composé à terme de six experts en sciences des données, ce Lab explore un large éventail de techniques telles que :

- le deep learning pour la prévision immédiate de nébulosité,
- l'hybridation machine learning / algorithmie classique pour la génération de bulletins automatiques,
- le data engineering pour la production d'indicateurs de risque météorologique à partir de la prévision d'ensemble,
- le croisement de données pour prévoir les interventions de pompiers sur événement météo à moyenne échéance.

Ce Lab IA s'appuiera dans un premier temps sur les machines GPU existantes, ainsi que sur un cluster GPU qui sera acquis dans le cadre du projet.

Le projet Deep4Cast

Mickaël Zamo - DIROP/COMPAS et Gregory Flandin – IRT Saint Exupéry

L'intelligence artificielle, au sens d'apprentissage machine, est utilisée depuis des années en Météorologie (dont Météo-France) sur des données climatologiques et météorologiques, notamment pour le post-traitement des modèles de prévisions numériques (adaptation statistique).

L'apprentissage profond ne fait pas partie des méthodes d'apprentissage machine couramment utilisées, à cause de sa relative nouveauté et du manque de recul de son intérêt sur les données météo.

Le but du projet Deep4Cast, financé par la fondation STAE, en partenariat avec l'IRT Saint Exupéry, le CERFACS et l'IRIT, est d'étudier l'intérêt de l'apprentissage profond pour la prévision de divers phénomènes météorologiques (précipitations, couverture nuageuse, téléconnexions) et à différentes échéances de prévision (de la prévision immédiate à la prévision saisonnière). L'apprentissage profond sera comparé à d'autres méthodes d'apprentissage machine, déjà éprouvées, afin d'étudier sa pertinence en tant que solution opérationnelle.

Lors de cette intervention, nous présenterons le cadre du projet Deep4Cast, ses problématiques pratiques et scientifiques et son état d'avancement.

Le chantier STAE DataNoos

Christophe Baehr - CNRM/GMEI

Nous présentons l'action DataNoos où le CNRM est fortement impliqué et les travaux qui en découlent. Les méthodes liées à l'IA reposent sur la structuration des jeux de données mais aussi sur l'interdisciplinarité. L'enjeu de DataNoos est de confronter les expériences des laboratoires toulousains sur la production, la gestion et la valorisation des données de la recherche. Trois cas d'usage ont été définis (Météo, Avion, Santé & environnement). Le CNRM, avec la MSHS-T, porte celui sur la génération de bulletins météorologiques orientés par les usages et l'expérience des bénéficiaires. Ces interactions nous ont permis de répondre à un AO de l'ANR avec la proposition « Semantics for Fair, usage-oriented semantic enrichment of open data » s'articulant autour de la donnée météorologique. Enfin nous parlerons des actions de réflexions et de formations (APSEM2018, ADO2019,...) autour de la structuration de la donnée dans un monde FAIR.

DONUT : optimisation de missions aéronautiques et spatiales sous incertitudes météorologiques

Florent Teichteil-Koenigsbuch - Airbus AI Research

Le projet DONUT vise à analyser l'impact des incertitudes pouvant intervenir en cours de mission (vol d'un avion commercial, observation de la Terre par un satellite) sur le plan de mission, et à optimiser ce dernier afin d'être robuste vis-à-vis des incertitudes. Les événements météorologiques sont parmi les impacts prédominants sur les coûts d'opération et les retards, que nous pouvons minimiser ou statistiquement contraindre à l'aide de techniques de décision séquentielle dans l'incertain et de modèles météorologiques probabilistes. Nous présenterons les cas d'études retenus dans le projet DONUT, la logique d'optimisation choisie ainsi que les performances obtenues au regard des technologies actuelles.

Retours d'expérience

Tour d'horizon des activités d'intelligence artificielle au CERFACS

Corentin Lapeyre – CERFACS

Les progrès de l'intelligence artificielle investissent de très nombreux domaines, mais ces approches "orientées données" peinent à trouver leur place dans la description de problèmes physiques, traditionnellement "orientés modèles". Le Cerfacs a une expertise en calcul intensif et modélisation physique, et investit depuis 1 an plusieurs pistes d'hybridation de ces approches avec des techniques d'apprentissage machine au sein du groupe de travail Helios.

Un tour d'horizon des synergies explorées sera donné dans cette présentation, incluant des applications en analyse d'images, traitement de séries temporelles, modèles de sous-maille et assimilation de données.

Des réseaux de neurones pour l'exploitation de bases de données multi-sources : application à la télédétection et à l'assimilation de données

Nemesio Rodriguez – CESBIO

Des réseaux de neurones ont été implémentés pour différentes applications dans le cadre de la mission SMOS (Soil Moisture and Ocean Salinity) de l'Agence Spatiale Européenne (ESA) et du Centre National d'Études Spatiales (CNES). Dans cette présentation je discuterai :

L'intérêt de ce type d'approches statistiques pour exploiter la synergie multi-capteur pour la mesure de l'humidité du sol et de la biomasse, par exemple pour la construction de longues séries temporelles

L'architecture du processeur d'humidité du sol SMOS en temps quasi-réel de l'ESA

L'utilisation des réseaux de neurones comme opérateur d'observation dans un contexte d'assimilation de données, approche qui a donné lieu à l'assimilation opérationnelle de données SMOS au CEPMMT

Traitement de profils verticaux issus du modèle AROME par une méthode d'apprentissage – application à la prévision du plafond nuageux aéronautique

Pierre Crispel - DSM/AERO

Nous présentons un exemple d'utilisation de méthodes d'apprentissage pour élaborer un diagnostic de plafond aéronautique nuageux à partir de profils verticaux de basse couche de différents paramètres issus du modèle AROME. Ces travaux s'inscrivent dans le projet IniTAF (Initialisation de TAF – Terminal Aerodrom Forecast), pour lequel ce sont des passages de seuils que nous souhaitons prévoir : 30, 60, 150, 300, 450, 1500, 5000 m. Deux diagnostics sont comparés : un diagnostic de "référence" basé sur la nébulosité 3d d'AROME utilisant la définition OACI du plafond aéronautique, un second diagnostic basé sur une méthode d'ensemble d'arbres de décision "Light Gradient Boosting Method" adaptée au traitement de jeux d'apprentissage de grande dimensionnalité. La vérité terrain est donnée par 2 années d'observations METAR horaires (66 aéroports - $5 \cdot 10^5$ obs). Des possibilités d'utilisation de réseaux de neurones pour ce problème sont discutées en perspectives.

Filtrage des données de radars météorologiques par Deep Learning

Pierre Lepetit - LATMOS - CNRM/GMAP

La donnée radar est essentielle pour localiser les précipitations et en mesurer l'intensité. Au début de la chaîne de traitement, ces données contiennent en mélange des échos de précipitations et des signaux parasites comme les échos de sol et les échos de ciel clair. Ces signaux parasites biaisent l'évaluation des cumuls de précipitations et compliquent l'interprétation de l'image radar. Il est donc important de pouvoir les détecter (tâche de « segmentation sémantique ») et les soustraire (« restauration »).

Nous avons abordé ces problèmes de segmentation et de restauration d'image par Deep Learning. La méthode d'apprentissage développée est simple et objective : les données n'ont pas besoin d'être labellisées « à la main ». Dans un premier temps, la méthode sera décrite et située par rapport à un état de l'art en pleine évolution. Les premiers résultats, relatifs au radar météorologique de Trappes, seront présentés.

La chaîne de traitement implique aussi une conversion de la réflectivité en taux de précipitations. Nous terminerons l'exposé en discutant des possibilités offertes par le Deep Learning sur cette question.

La détection et la caractérisation automatique de la texture des précipitations

Yamina Hamidi - CNRM/GMAP, Raynaud Laure, Rottner Lucie et Arbogast Phillipe

Les modèles à haute résolution comme AROME et la prévision d'ensemble AROME arrivent à prévoir les précipitations d'une manière réaliste, mais des erreurs de positionnement, de chronologie ou d'intensité persistent.

Afin d'améliorer l'utilisation de ces prévisions une méthode de spatialisation a été utilisée. Elle consiste à extraire du champ prévu un signal de plus grande échelle que la maille du modèle et donc plus prévisible sous la forme d'« objets précipitants ». Un objet est défini comme une région dans laquelle la distribution des précipitations est homogène en intensité et/ou en variabilité spatiale (texture). Cette approche présente également l'avantage d'imiter la démarche des prévisionnistes qui effectuent un zonage du territoire « à la main ».

La détection des objets de pluie est actuellement effectué à partir de leur intensité : on caractérise les objets de pluies totales, modérées et fortes. Afin de fournir une description complète du temps sensible la texture de ces objets doit également être caractérisée, c'est-à-dire leur nature continue ou intermittente.

Cet exercice de classification de la texture est un cas d'application classique des méthodes de machine learning. On présentera les performances de deux classifieurs, la régression logistique et les forêts aléatoires ainsi que leur sensibilité aux prédicteurs et aux hyper-paramètres choisis.

Apprentissage profond et altimétrie satellitaire pour le suivi des eaux continentales.

Denis Blumstein – CNES

La gestion de la ressource en eau à l'échelle mondiale représente d'ores et déjà un défi majeur qui prendra de plus en plus d'ampleur dans les décennies à venir. Les élévations de la surface des rivières et des lacs font partie des variables climatiques essentielles, telles que définies par le Global Climate Observing System (GCOS). Traditionnellement, ces observations se font via des mesures in situ de cotes d'eau et de débit. Pour diverses raisons (financières, stratégiques, géopolitiques, ...), ces réseaux de stations in situ ont un échantillonnage spatio-temporel particulièrement hétérogène et ont tendance à diminuer depuis quelques décennies. Depuis deux décennies, de nombreuses équipes de recherche ont démontré que les mesures produites par les altimètres radar spatiaux pouvaient être utilisées pour compenser cette diminution.

Cependant l'utilisation de ces mesures nécessite des post-traitements délicats à cause du principe de mesure utilisé et de l'immense variété des situations géophysiques rencontrées. Cette activité consomme une trop grande quantité de temps ingénieur.

On présente ici une application de l'apprentissage profond (réseaux de type RESNET-50) à l'analyse de radargrammes en vue de fournir le niveau des plans d'eau survolés : le problème traité est donc un problème de régression plutôt que les problèmes de classification classiquement traités. Cette technique ouvre la voie à une automatisation du processus qui permettra d'exploiter tout le potentiel des radars altimétriques en orbite ou à venir.

Utilisation de l'IA pour remplacer tout ou partie des modèles atmosphériques

Aurélien Ribes - CNRM/GMGEC

Nous présentons deux utilisations de l'IA pour la modélisation de l'atmosphère, en remplacement de tout ou partie de modèles physiques. Il s'agit dans les deux cas de travaux de recherche, avec des résultats très partiels.

Dans le premier travail, nous cherchons à apprendre les processus physiques sous-maille (les "paramétrisations physiques") du modèle global ARPEGE-Climat, incluant le rayonnement, la convection, la turbulence, etc. Cela équivaut à prédire, sur une colonne atmosphérique, et à l'échelle du pas de temps du modèle, les tendances de température et d'humidité à partir de données d'entrée pertinentes, que nous avons sélectionnées parmi les profils de température, d'humidité, de vent zonal ainsi que les flux de chaleurs, le rayonnement, etc. L'échantillon d'apprentissage est issu d'une simulation d'ARPEGE-Climat (éventuellement à plus haute résolution). Différents algorithmes de machine learning sont testés, parmi lesquels réseaux de neurones et forêts aléatoires.

Dans le second travail, nous cherchons à émuler un modèle de climat régional (RCM). Les RCM sont des modèles à aire limitée qui permettent de faire des simulations à haute résolution à partir de conditions au bord, issues d'un modèle global. Le coût de calcul élevé de ces modèles ne permet pas de couvrir une large gamme d'incertitude / scénarios. Nous cherchons donc à émuler leur sorties (ex: les variables physiques en un point du domaine) à partir des entrées (les conditions au bord), via diverses méthodes d'apprentissages statistiques.

Représentations et analyses sémantiques en Traitement Automatique des Langues

Philippe Muller – UPS/IRIT

Le traitement automatique des langues (TAL) a pour objectif de modéliser les phénomènes linguistiques, en vue d'automatiser la compréhension et la génération de messages textuels. Il se décline en une multitude de problèmes, motivés par des applications particulières, de plus en plus nombreuses, ou par la recherche de représentations intermédiaires "intéressantes". Les réseaux de neurones sont devenus un outil majeur en TAL en fournissant des représentations intermédiaires utiles pour modéliser le sens des mots, de phrases ou de textes en langage naturel, permettant des avancées importantes. Nous présenterons les problèmes principaux et les modèles récents de deep learning associés.