Chapitre 1

Cadre général

L’apparition de la photographie et l’essor de l’aviation puis la mise en place de satellite d’observation de la terre ont permis le développement d’ensemble de techniques utilisées pour déterminer à distance les propriétés d'objets naturels ou artificiels à partir des rayonnements qu'ils émettent ou réfléchissent, c’est la télédétection. Ce vaste champ d’étude a plusieurs applications très pratiques notamment en météorologie, en reconnaissance militaire, gestion des ressources agricoles et forestières, cartographie etc (Maya Nand Jha et al., 2019). Tout une branche de ces méthodes a pour objet l’étude de la propriété de certains milieu ou objets à partir de caractéristiques physiques mesurables notamment le rayonnement pour les identifier et donc faire de la classification. C’est donc assez logiquement que l'utilisation de méthodes statistiques et autres modèles d’apprentissages comme les arbres de décision a fait son essor notamment pour la caractérisation des cultures ou la cartographie (Gbodjo et al. (2020), Kumar et al. (2018), Zafari et al. (2017) et Gislason et al. (2005)).

Dans le cadre de notre projet, nous avons comparé un modèle de forêt aléatoire à un modèle de forêt aléatoire avec bootstrap aggregating pour la classification orienté pixel d'images GeoTif de la zone de berlin alentour selon. Le bootstrap aggregating (bagging) étant un algorithm d'apprentissage ensembliste permettant d'améliorer la stabilité et la précision des algorithmes d'apprentissage automatique en réduisant la variance et permettant d'éviter le surapprentissage. On cherche donc à déterminer si l'algorithme Bagging combiné au modèle classique apporte une amélioration significative dans l'utilisation de ces méthodes dans le domaine de la classification d'image satellites.

Une image contenant diagramme

Description générée automatiquement

* 1. Objectif projet

Ce projet s’inscrit dans le cadre du cours d’Apprentissage Statistique et IA et s’inspire d’un module du séminaire Introduction to Remote Sensing destiné aux étudiants en géographie de la Humboldt-Universität zu Berlin. L’objectif générale est d’améliorer un modèle de Random Forêt pour classer des images staellites en le combinant avec l’agorithm Bagging. Il s’agit plus spécifiquement d’implémenter un modèle de Random Forêt sur des images satellites de la zone de berlin et alentour pour classer les element sur les images pixel par pixel selon s’il sont : Prairies/herbacées(Grassland/Herbaceous) , Végétation ligneuse à feuilles larges (Broadleafed woody vegetation), Végétation ligneuse de conifères (Coniferous woody vegetation), Construit (Built-up), L'eau (Water), Terres cultivées (Cropland) à partir des bands spectraux des images puis d’appliquer l’algorithm bagging au model et comparer.

Chapitre 2

Contexte Méthodologique

2.1 État de l’Art

La littérature sur les techniques et méthodes statistiques utilisées en télédétection est bien fournie. En particulier, l'utilisation d'arbres de décision a été largement étudiée. Elles sont très utilisées pour extraire des informations à partir des images de télédétection, tandis que la méthode Random Forest est une méthode d'ensemble efficace pour la classification de la couverture terrestre. Pour la classification orienté apprentissage machine, on distingue principalement deux approches complémentaires celle basée sur les pixels avec un classificateur Random Forest et celle orienté objets avec une méthode de segmentation d'objet basée sur les relations hiérarchiques (Bui et Mucsi, 2020). Il apparait que les performances de modèles Random Forest comparées à d'autres méthodes de classification donnent des résultats comparables ou meilleurs (Gislason et al. 2006). Mais même si l’idée de réunir plusieurs arbres de décision en une forêt aléatoire permet d’obtenir une stabilité du modèle, dans une perspective de prospectives, le contrôle de surapprentissages et la précision du modèle est un impératif. Breiman (1996) a proposé à cet effet, la méthode de bootstrap aggregating (bagging) pour la prédiction à l'aide de plusieurs modèles de prédiction. Le bagging est une technique de combinaison de modèles qui utilise un échantillonnage avec remplacement pour produire plusieurs sous-ensembles de données d'entraînement pour chaque modèle. Mise en perspectives avec d’autres méthode d'ensembles d'arbres de décision (le Boosting et la Randomisation), le Bagging apparait efficace pour réduire la variance des prédictions (Dieterich, 2000).

Chapitre 3

Modélisation

3.1 L’état de l’art

3.2 Les modèles

3.2.1 Modèle DT

3.2.2 Modèle DT Bagging

3.3 Le jeux de données

<https://github.com/Oniheii/Teledetection-Bootstrap-Aggregating-Random-Forest>

<https://scihub.copernicus.eu/dhus/#/home>

3.4 Résultats

Chapitre 4

Conclusion et perspectives

Il s’avère que la forêt aléatoire est déjà particulièrement performant pour cette classification plus de 90% de précision ce qui peut s’expliquer par la nature de la règle d’étiquetage des pixels basée sur des seuils pour les profiles spectraux c’est à les valeurs de matrices de pixel (Raster).