

ALGO : RANDOM FOREST

Etape 1

Liens :

[RandomForestClassifier](#)

[RandomForestRegressor](#)

[Wikipédia](#)

Nous avons choisi de travailler sur un algorithme de type Random Forest (Forêt aléatoire).

Cet algorithme est applicable à des problèmes de classification (RandomForestClassifier) comme à des problèmes de régression (RandomForestRegressor).

Pourquoi nous avons choisis cet algorithme :

Ayant étudié les structures arborescentes durant notre cursus, et sachant que c'est un outil très utilisé dans beaucoup de domaines, nous étions d'abord intéressés par les arbres de décision.

Mais nous avons ensuite découvert les algorithmes de type forêts aléatoires, qui nous ont paru plus complexes et enrichissants à étudier que les seuls arbres de décision. Ils sont d'ailleurs apparemment plus fiables et plus précis que ces derniers.

De plus, ces algorithmes ont l'air d'être assez souvent utilisés et reconnus (ils ont par exemple été utilisés pour la reconnaissance des mouvements du Kinect sur la console Xbox 360). C'est pourquoi les étudier pour mieux les comprendre nous paraît important, car nous y serons sûrement de nouveau confrontés.

Scénarios possibles :

- Reconnaissance d'images de tous types : d'objets, de plantes, de visages ou encore de mouvements (à été utilisé par exemple pour la reconnaissance de mouvements sur la Kinect de la console Xbox 360).

- Identification de sons, par exemple pour une reconnaissance vocale, ou pour des applications musicales (reconnaisances de notes, d'accords, de tempos, d'instruments...).
- Reconnaissance de textes, de caractères, par exemple pour catégoriser des écrits ou dans applications de type chatbots.
- Prédiction de valeurs numériques, comme un nombre de places disponibles en réanimation dans les hôpitaux, des chiffres sur le climat, sur la qualité de l'air, des évolutions de taux financiers... et bien d'autres (Utilisant donc des données quantitatives).
- Prédiction de comportements. Par exemple, des prévisions de comportements météorologiques, ou encore dans un but commercial, comme des prédictions de comportement des clients (Utilisant des données quantitatives comme qualitatives).
- Identification selon des données qualitatives, par exemple des goûts musicaux ou cinématographiques, dans un but de recommandations. Identification de maladies selon des états, des symptômes... .

<https://www.kaggle.com/rishidamarla/films-dataset>

<https://www.kaggle.com/mazenramadan/imdb-most-popular-films-and-series>

Jeux de test

- classification
- champignons, maladie,
- films : recommandation personnelle, succes film ou autres
- [image satellites](#)
- attributs éruptions solaires (prédire une éruption)
- prédire qualité de l'air bof

Classification :

- cancer de la peau
<https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000>
- classification étoiles, galaxies, quasar
<https://www.kaggle.com/fedesoriano/stellar-classification-dataset-sdss17>
- animaux
<https://www.kaggle.com/sharansmenon/animals141>
- cancer du sein
<https://www.kaggle.com/yasserh/breast-cancer-dataset>
- tumor brain
<https://www.kaggle.com/jillanisofttech/brain-tumor>
- prediction maladie symptômes (mis dans classification mais ça doit être de la prédiction je pense)
<https://www.kaggle.com/itachi9604/disease-symptom-description-dataset>
- langage des signes (reconnaitre ce qui est dit)ouai
<https://www.kaggle.com/anirudhchauhan/sign-language-recognition-alphabets>

<https://www.quantmetry.com/blog/classification-et-desequilibre-de-classes/>

