Forêt Aléatoire et Clustering Prédictif en Chemoinformatique

Sébastien Ramel[1], Simon Bernard[1], Laurent Heutte[1]

[1]Université de Rouen Normandie, LITIS, 76000 Rouen
$2^{ème}$ *réunion du projet SCHISM, GREYC, Caen*

13/12/2021

Projet SCHISM [1]

Contexte

Application chémoinformatique concernant l'analyse des relations entre la structure d'une molécule et son activité inhibitrice d'une protéine cible d'intérêt (responsable d'une maladie).

Objectif double

1. apprentissage automatique pour modéliser la relation entre structure moléculaire et activité
2. exploration de données pour expliquer cette relation et permettre l'interaction avec un expert.

Proposition

Baser la modélisation sur l'apprentissage de similarités entre molécules et l'explicabilité sur l'analyse des liens entre des sous-structures et l'activité.

1. Albrecht Zimmermann. *PROJET SCHISM*. 2021. url : https://schism.greyc.fr/.

Jeu de données considéré [2]

- 1485 molécules décrites par $F = 112048$ caractéristiques binaires dont l'activité (connue) concerne l'inhibition de la tyrosine kinase (responsable de la leucémie).
- Nombre de caractéristiques réduit à $F = 15129$, grâce à la suppression de caractéristiques redondantes.

$\Rightarrow$ Problème impliquant des données à (très) grande dimension

- $\Rightarrow$ distances euclidiennes non pertinentes,
- $\Rightarrow$ apprentissage de modèle difficile,
- $\Rightarrow$ risques de sur-apprentissage fort,
- $\Rightarrow$ très grand nombre de sous-structures candidates.

2. Albrecht Zimmermann. *DONNEES SCHISM*. 2021. url : https://unicloud.unicaen.fr/index.php/s/sS6WqQkGZpfDdEJ?path=%5C%2Fschism.

Approche

- Hypothèse : plus des molécules sont structurellement similaires, plus leurs activités sont similaires.
- Décrire, prédire et expliquer l'activité de molécules

  Décrire la structure sous-jacente de l'activité (*e.g.* sous familles, cliffs) par des clusters de molécules similaires,

  Prédire l'activité d'une molécule de test d'après son cluster le plus proche,

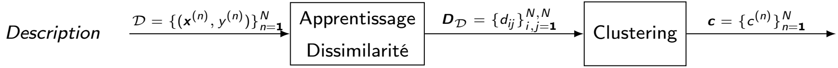
  Expliquer la prédiction via les contributions apportées par chacune des caractéristiques pour former le cluster.

Apprentissage de dissimilarité avec des forêts aléatoires

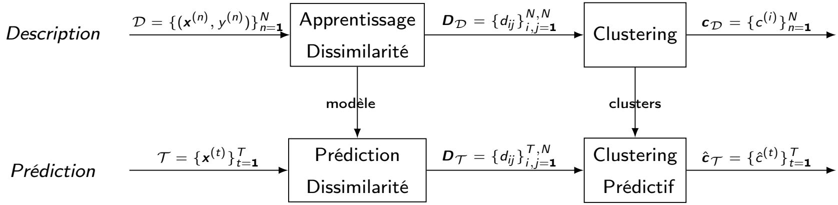
Algorithme des forêts aléatoires

- intègre une mesure de similarité basée sur les caractéristiques et l'appartenance aux classes (activité),
- robuste aux données à dimension élevée,
- fournit des outils d'analyse et d'interprétabilité.

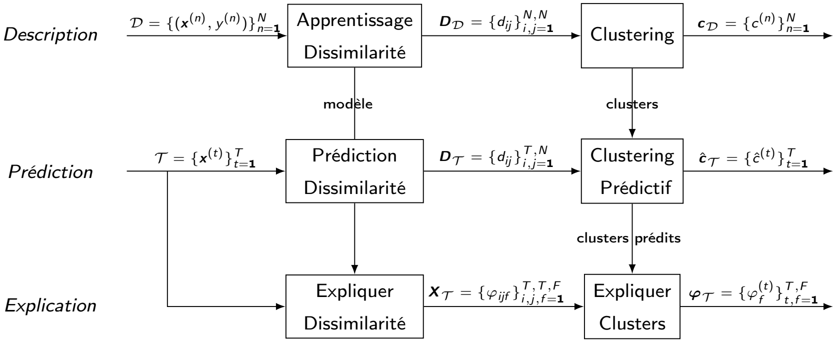
Clustering supervisé



Clustering prédictif



Clustering explicable



Plan

1. Forêt aléatoire
   - Arbre décisionnel
   - Ensemble d'arbres
   - Mesure de proximité
2. Expériences
   - Clustering prédictif
   - Paramètres impactants
   - Résultats
3. Explicabilités
   - Méthodes
   - Perspectives

Plan

1. Forêt aléatoire
   - Arbre décisionnel
   - Ensemble d'arbres
   - Mesure de proximité
2. Expériences
   - Clustering prédictif
   - Paramètres impactants
   - Résultats
3. Explicabilités
   - Méthodes
   - Perspectives

Structure d'un arbre décisionnel

- Une arbre h modélise une relation de dépendance

$$h : \mathcal{X} \rightarrow \mathcal{Y}$$

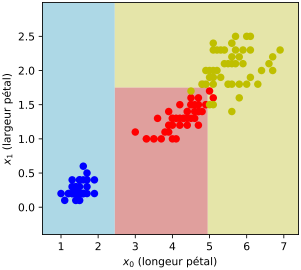
  entre une variable cible Y et des caractéristiques $\boldsymbol{x} = (x_1, \ldots, x_F) \in \mathcal{X}$
- h est composé d'un ensemble $\mathcal{N}$ de nœuds $t \in \mathcal{N}$, hiérarchisés de la racine aux feuilles.
- Chaque nœud interne t (*i.e.* $\neq$ feuilles) possède une question sur sa caractéristique f_t associée

$$q_t : x_{f_t} \leq d \;?,$$

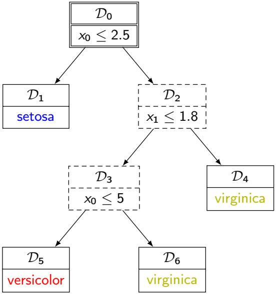
  qui divise ses instances $\mathcal{D}_t \subseteq \mathcal{D}$ dans 2 nœuds enfant : t_l et t_r
- Division de t choisie en maximisant la réduction de l'impureté (*e.g.* entropie de Shannon) de la répartition des classes de $\mathcal{D}_t$

Exemple sur une base de données jouet (Iris)

avec $\mathcal{X} = \mathbb{R}^2$ et $\mathcal{Y} = \{\texttt{setosa}, \texttt{versicolor}, \texttt{virginica}\}$



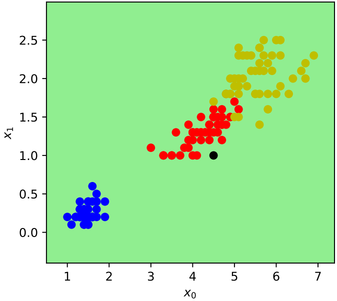
(a) Partition donnée par les feuilles d'un arbre entrainé h



(b) Ensemble $\mathcal{N}$ des nœuds de h : racine (=), internes (--) et feuilles (–)

Prédiction de $\boldsymbol{x}^{(tst)} = (4.5, 1)$

Instances $\mathcal{D}_0$ du nœud racine $t = 0$



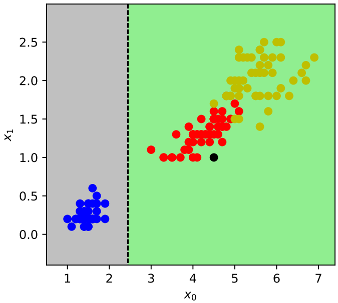
(a) Sous ensemble (■) atteint par $\boldsymbol{x}^{(tst)}$ (●)

$\mathcal{D}_0$; ●
$x_0 \leq 2.5$

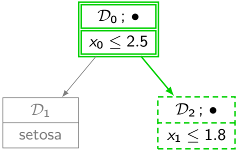
(b) Chemin $\mathcal{C}^{(tst)} = \{0\}$ suivi par $\boldsymbol{x}^{(tst)}$ (●).

Prédiction de $\boldsymbol{x}^{(tst)} = (4.5, 1)$

Décision du nœud racine $t = 0$



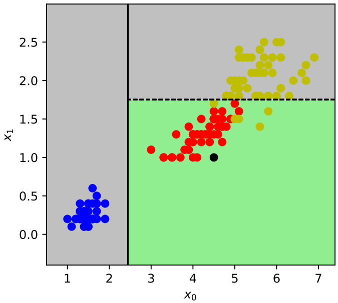
(a) Sous ensemble (■) atteint par $\boldsymbol{x}^{(tst)}$ (●)



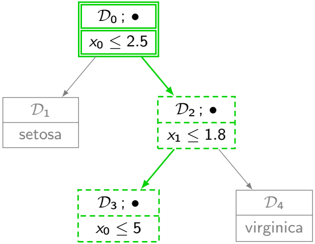
(b) Chemin $\mathcal{C}^{(tst)} = \{\{0\}, \{2\}\}$ suivi par $\boldsymbol{x}^{(tst)}$ (●).

Prédiction de $\boldsymbol{x}^{(tst)} = (4.5, 1)$

Décision du nœud $t = 2$



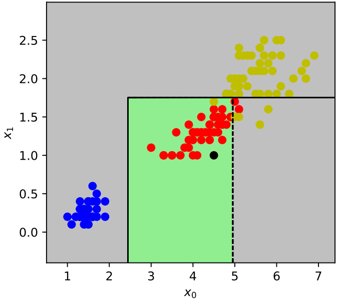
(a) Sous ensemble (■) atteint par $\boldsymbol{x}^{(tst)}$ (●)



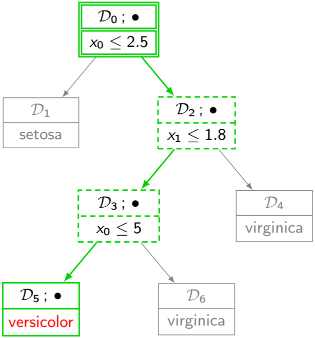
(b) Chemin $\mathcal{C}^{(tst)} = \{\{0\}, \{2\}, \{3\}\}$ suivi par $\boldsymbol{x}^{(tst)}$ (●).

Prédiction de $\boldsymbol{x}^{(tst)} = (4.5, 1)$

Décision du nœud $t = 3$



(a) Sous ensemble (■) atteint par $\boldsymbol{x}^{(tst)}$ (●)



(b) Chemin $\mathcal{C}^{(tst)} = \{\{0\}, \{2\}, \{3\}, \{5\}\}$ de $\boldsymbol{x}^{(tst)}$ (●). $\hat{y}^{(tst)} = \texttt{versicolor}$

Forêt aléatoire (FA) [3]

- L'arbre h est simple et prédictif mais manque de stabilité.
- La FA $= \{h_1, \ldots, h_{N_{\text{arbre}}}\}$, corrige ce défaut en agrégeant les décisions de N_{arbre} arbres diversifiés.
- Diversité produite par deux processus de "randomisation" dans la construction des arbres

  Bagging : échantillonnage aléatoire (avec remise) des données d'apprentissage $\mathcal{D}$,

  Random Feature Selection : échantillonnage aléatoire des F caractéristiques avant d'identifier la caractéristique associée à la division de chaque nœud.

3. Leo Breiman. "Random Forests". *45* (2001), p. 5-32.

Mesure de (dis)similarité d'un arbre h

- Soient $(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) \in \mathcal{X}^2$, deux instances dont on souhaite connaitre leur ressemblance.

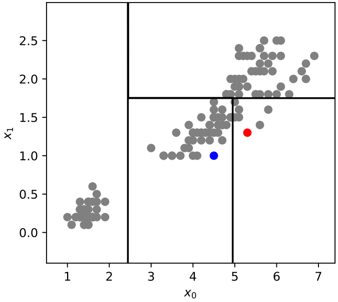
Similarité $s(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ mesurée par un arbre

$$s(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = \begin{cases} 1 & \text{si } \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \text{ parcourent le même chemin,} \\ 0 & \text{sinon.} \end{cases}$$

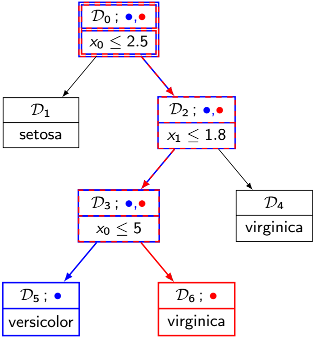
- $\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}$ sont similaires si elles sont suffisamment rapprochées pour atteindre la même feuille et donc probablement de classe identique.
- La mesure de dissimilarité $d(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ est donnée par

$$d(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = 1 - s(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}).$$

Similarité de $\boldsymbol{x}^{(tst)} = (4.5, 1)$ et $\boldsymbol{x}^{(u)} = (5.3, 1.3)$



(a) Feuille atteinte par chacune des instances $\boldsymbol{x}^{(tst)}$ (●), $\boldsymbol{x}^{(u)}$ (●)



(b) Chemin suivi par chacune des instances. $d(●, ●) = 1$.

Mesure de dissimilarité d'une FA

Dissimilarité $d_{FA}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ mesurée par une FA

$$d_{FA}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = \frac{1}{N_{arbre}} \sum_{m=1}^{N_{arbre}} d_m(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}),$$

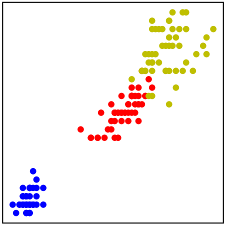
avec $d_m(\cdot, \cdot)$, la dissimilarité de l'arbre $h_m \in \text{FA} = \{h_1, \ldots, h_{N_{arbre}}\}$.

- Matrice de dissimilarité donnée par

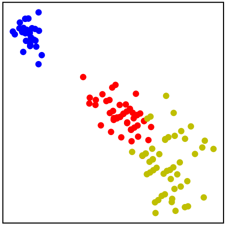
$$\boldsymbol{D}_{FA} = \begin{bmatrix} d_{FA}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(1)}) & \cdots & d_{FA}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(N)}) \\ \vdots & \ddots & \vdots \\ d_{FA}(\boldsymbol{x}^{(N)}, \boldsymbol{x}^{(1)}) & \cdots & d_{FA}(\boldsymbol{x}^{(N)}, \boldsymbol{x}^{(N)}) \end{bmatrix}$$

Positionnement 2D des dissimilarités $\boldsymbol{D}_{FA}$

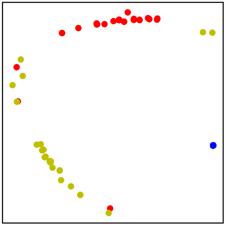
Jeu de données Iris



(a) Données originales.



(b) Distances euclidiennes.



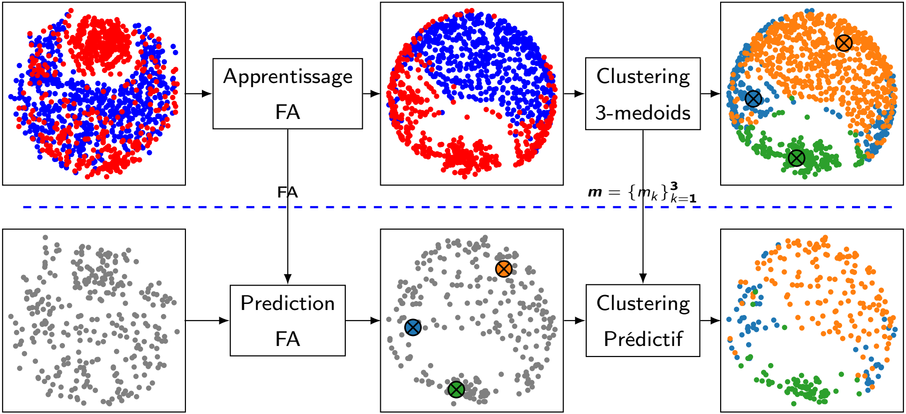
(c) Dissimilarités $\boldsymbol{D}_{FA}$ (supervisée).

Plan

1. Forêt aléatoire
   - Arbre décisionnel
   - Ensemble d'arbres
   - Mesure de proximité
2. Expériences
   - Clustering prédictif
   - Paramètres impactants
   - Résultats
3. Explicabilités
   - Méthodes
   - Perspectives

K-medoids basé sur D_{FA}

Données du projet SCHISM



Mesures de performance[4] sur base de test $\mathcal{T}$

Validation croisée (10 blocs)

Adjusted Random Index (ARI) Taux de paires d'instance correctement regroupées sachant la partition donnée par leurs classes,

Adjusted Mutual Information (AMI) Information mutuelle (ajustée par chance) de la partition prédite avec celle formée par les classes

Normalized Mutual Information (NMI) Information mutuelle (normalisée) de la partition prédite avec celle formée par les classes

Silhouette score (SIL) Mesure la capacité des clusters à regrouper des instances similaires et à dissocier des instances différentes

4. Dongkuan Xu et Yingjie Tian. "A Comprehensive Survey of Clustering Algorithms". *Annals of Data Science* 2 (août 2015), p. 165-193.

Étude des paramètres impactants des FA

Taille ($\mathrm{maxFeatures}$) du vecteur de caractéristique échantillonné

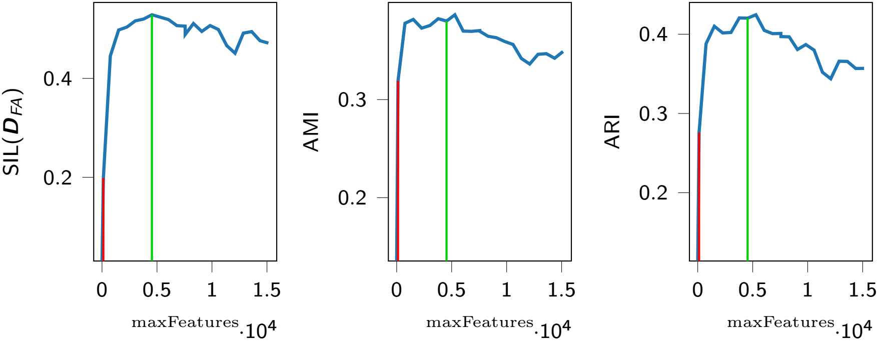
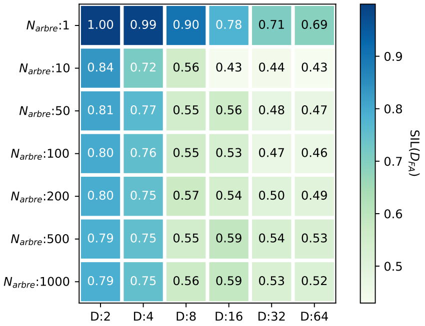


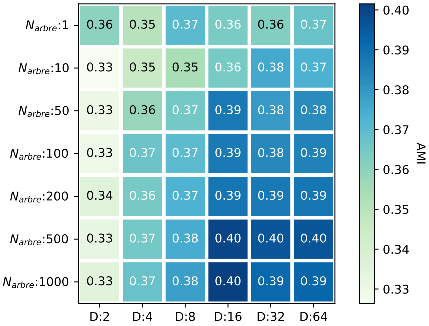
Figure – Performances (—) en fonction du paramètre $\mathrm{maxFeatures}$. Valeur par défaut : $\mathrm{maxFeatures} = \sqrt{F} = 123$ (—). Valeur choisie : $\mathrm{maxFeatures} = 4536$ (—).

Étude des paramètres impactants des FA

Nombre d'arbre (N_{arbre}) et profondeur limite (D)



(a) Score Silhouette basé sur $\boldsymbol{D}_{FA}$



(b) Adjusted Mutual Information

Performances externes

Basées sur les étiquettes de classe de $\mathcal{T}$

Type Clust.	Non Supervisé		Supervisé	
Diss. / Perf.	Eucl.	Jacc.	$\boldsymbol{D}_{FA}$	MLP+Eucl
ARI	$.142 \pm .063$	$.105 \pm .021$	$.420 \pm .070$	
AMI	$.219 \pm .030$	$.140 \pm .035$	$.380 \pm .049$	
NMI	$.240 \pm .027$	$.156 \pm .034$	$.396 \pm .047$	$.299 \pm .024$

Table – Performances externes des méthodes de clustering basées sur différentes mesures de dissimilarité.

Performances internes

Basées seulement sur les caractéristiques de $\mathcal{T}$

Type Clust.	Non Supervisé		Supervisé	
Diss. / SIL	Eucl.	Jacc.	$\boldsymbol{D}_{FA}$	MLP+Eucl
Eucl.	$.047 \pm .031$	$-.009 \pm .013$	$-.098 \pm .025$	$.476 \pm .014$
Jacc.	$.053 \pm .025$	$.054 \pm .010$	$.038 \pm .008$	
$\boldsymbol{D}_{FA}$	$-.057 \pm .117$	$-.0776 \pm .030$	$.529 \pm .049$	

Table – Scores Silhouette des méthodes de clustering évalués dans différents espaces.

Plan

1 Forêt aléatoire
- Arbre décisionnel
- Ensemble d'arbres
- Mesure de proximité

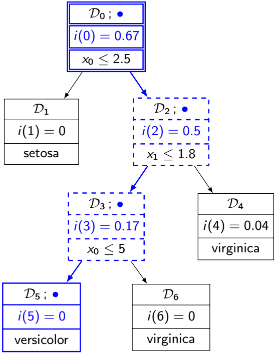
2 Expériences
- Clustering prédictif
- Paramètres impactants
- Résultats

3 Explicabilités
- Méthodes
- Perspectives

Importance statique des caractéristiques

Mean Decrease in Impurity (MDI) d'une instance $\boldsymbol{x}^{(i)}$

- Mesure l'importance moyenne d'une caractéristique X_j dans la prédiction des instances $(\boldsymbol{x}^{(i)}, y^{(i)}) \in \mathcal{D}$.
- Chaque instance $(\boldsymbol{x}^{(i)}, y^{(i)})$ suit un chemin $\mathcal{C}^{(i)}$ composé d'une série de nœuds dotés d'un test sur une caractéristique.
- Les gains en réduction d'impureté des nœuds basés sur X_j rencontrés sur $\mathcal{C}^{(i)}$, sont accumulés.
- Les gains accumulés sont moyennés sur l'ensemble des arbres $h \in FA$.



(a) Chemin $\mathcal{C}^{(t)}$ suivi par $\boldsymbol{x}(t)$ (•). Impureté $i(t)$ des nœuds de h.

Importance statique des caractéristiques

Mean Decrease in Impurity (MDI)

- Proportion d'instance atteignant t à partir de v

$$p_v(t) = \frac{|\{(\boldsymbol{x}^{(i)}, y^{(i)}) \in \mathcal{D}_t\}|}{|\{(\boldsymbol{x}^{(i)}, y^{(i)}) \in \mathcal{D}_v\}|}$$

- Gain en réduction d'impureté de t

$$\Delta(t) = i(t) - p_t(t_l)i(t_l) - p_t(t_r)i(t_r)$$

- MDI d'une variable X_j

$$\mathrm{MDI}(X_j) = \frac{1}{N_{arbre}} \sum_{m=1}^{N_{arbre}} \sum_{t=1}^{|\mathcal{N}_m|} \mathbb{1}(f_t = j) p_0(t) \Delta(t)$$

Importance statique des caractéristiques en clustering

MDI local à un cluster k

- On note $p_v(t,k)$ la proportion d'instance du cluster k atteignant le nœud t à partir de v

$$p_v(t,k) = \frac{|\{(\boldsymbol{x}^{(i)}, y^{(i)}) \in \mathcal{D}_t | c^{(i)} = k\}|}{|\{(\boldsymbol{x}^{(i)}, y^{(i)}) \in \mathcal{D}_v | c^{(i)} = k\}|}$$

- MDI d'une variable X_j local au cluster k

$$\mathrm{MDI}(X_j, k) = \frac{1}{N_{arbre}} \sum_{m=1}^{N_{arbre}} \sum_{t=1}^{|\mathcal{N}_m|} \mathbb{1}(f_t = j) p_0(t,k) \Delta(t)$$

MDI des groupes de passagers du Titanic

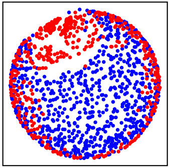
Jeu de données

- Survivants du Titanic décrits par 7 caractéristiques

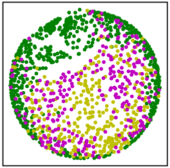
pclass	Classe du ticket
sex	Sexe du passager
age	Age du passager
sibsp	# de frères et sœurs / conjoints
parch	# de parents / enfants
fare	Tarif passager
embarked	Port d'embarquation

- Ajout de 2 caractéristiques non informatives

rand_num	valeurs numériques
rand_cat	valeurs catégoriques

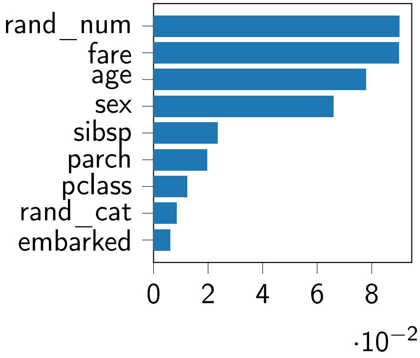


(a) Passagers : survivants (●), morts (●)

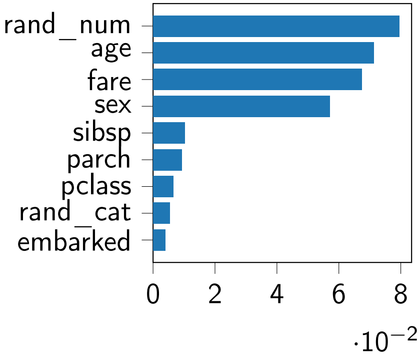


(b) Clusters : 1 (●), 2 (●) et 3 (●)

MDI des groupes de passagers du Titanic



(a) MDI local au cluster 1



(b) MDI local au cluster 2

Autre méthodes d'explicabilité spécifiques aux FA

- Sur le même principe que MDI (*i.e.*, exploration des chemins parcourus par des instances), 2 autres méthodes

  Mean Decrease in Accuracy (MDA) Pour chaque X_j, permuter aléatoirement ses valeurs dans $\mathcal{T}$ et calculer la perte d'accuracy. Moyenner ces pertes sur l'ensemble des arbres de la FA.

  SHAP method for tree (Tree SHAP) [5] Explicabilité dynamique SHAP des prédictions d'une FA. Tire parti d'une complexité polynomiale

$$\mathcal{O}(N_{arbre} \cdot L \cdot D^2), \quad L = \text{nb max feuille}.$$

5. Scott Lundberg et al. "From Local Explanations to Global Understanding with Explainable AI for Trees". *Nature Machine Intelligence* (2020), p. 56-67.

Méthodes SHAP [6]

- Dans un jeu coopératif, la valeur de Shapley retourne une répartition équitable du gain $v(N)$ généré par n joueurs d'une coalition N.
- Valeur Shapley d'une caractéristique x_f dans la prédiction $\hat{y}(\boldsymbol{x})$

$$\varphi_f(\hat{y}) = \sum_{S \subseteq \mathcal{F} \setminus \{f\}} \frac{|S|!(F - |S| - 1)!}{F!} \left(\hat{y}(\boldsymbol{x}_{S \cup \{f\}}) - \hat{y}(\boldsymbol{x}_S)\right)$$

où $\mathcal{F} = \{1, \ldots, F\} \ni f$

- La méthode SHAP suppose $\hat{y}(\boldsymbol{x}_S) = \mathbb{E}[\hat{y}(\boldsymbol{x}) | \boldsymbol{x}_S]$.

6. Scott Lundberg et Su-In Lee. “A unified approach to interpreting model predictions”. (2017).

Explication dynamique du clustering via Tree SHAP

Expliquer les dissimilarités $\boldsymbol{D}_{FA}$ puis le clustering

- Dans cette proposition, la cible de l'explication n'est plus la prédiction $\hat{y}(\boldsymbol{x}^{(i)})$, mais la similarité $s_{ij} = s(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ de 2 instances d'intérêt

$$\varphi_f(s_{ij}) = \sum_{S \subseteq \mathcal{F} \setminus \{f\}} \frac{|S|!(F - |S| - 1)!}{F!} \left(s(\boldsymbol{x}^{(i)}_{S \cup \{f\}}, \boldsymbol{x}^{(j)}_{S \cup \{f\}}) - s(\boldsymbol{x}^{(i)}_{S}, \boldsymbol{x}^{(j)}_{S}) \right)$$

- Explication de l'assignation d'une instance de test $\boldsymbol{x}^{(tst)}$ au cluster k le plus proche

$$\boldsymbol{\varphi}^{(tst)} = (\varphi_1(s_{tst,k}), \ldots, \varphi_F(s_{tst,k}))$$

Conclusions

Synthèse

- Décrire la structure de l'activité de molécules en utilisant la mesure (supervisée) de (dis)similarité d'une FA *i.e.* basée sur : proximité et classe des molécules.
- Prédire le cluster de molécules de test étant donné la description inférée à l'apprentissage.
- Proposition d'une variante à Tree SHAP pour expliquer les dissimilarités d'une FA et par la suite l'assignation des clusters : Similarity Tree SHAP.

Suites

- Implémenter Similarity Tree SHAP avec complexité polynomial (actuellement $\mathcal{O}(N_{arbre}, L, F, 2^F)$, rédhibitoire !).
- Trouver des exemples d'application en Chemoinformatique ou autre, avec une vérité terrain, pour valider la pertinence de cette approche.
- Déduire les autres outils d'analyse SHAP avec Similarity Tree SHAP en clustering.

Merci pour votre attention.

Importance des caractéristiques (classification)

Mean Decrease in Accuracy (MDA)

- Permuter aléatoirement les valeurs de X_j dans $\mathcal{T}$ pour former $\tilde{\mathbf{x}}_j \in \mathcal{X}$
- Pour chaque arbre h_m avec $m = 1, \ldots, N_{arbre}$
  - Calculer la nouvelle précision

$$\tilde{\mathrm{Acc}}(m) = \frac{1}{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{T}|} \mathbb{1}(h(\tilde{\mathbf{x}}_j^{(i)}) = y^{(i)})$$

  - Calculer la perte de précision

$$I(m) = \mathrm{Acc}(m) - \tilde{\mathrm{Acc}}(m)$$

- MDA d'une variable X_j

$$MDA(X_j) = \frac{1}{N_{arbre}} \sum_{m=1}^{N_{arbre}} I(m)$$

Importance des caractéristiques (clustering)

Mean Decrease in ARI local à un cluster k

- Permuter aléatoirement les valeurs de X_j dans $\mathcal{T}$ afin d'obtenir $\tilde{\boldsymbol{x}}_j^{(i)} \in \mathcal{X}$ pour $i = 1, \ldots, |\mathcal{T}|$
- Calculer les dissimilarités $d_{FA}(\tilde{\boldsymbol{x}}_j^{(i)}, \boldsymbol{x}^{(med(k))})$ de $\tilde{\boldsymbol{x}}_j^{(i)}$ avec les medoids $\boldsymbol{x}^{(med(k))}$ des clusters $k = 1, \ldots, K$
- Prédire les nouveaux clusters $\{\tilde{\hat{c}}_i\}_{i=1}^{|\mathcal{T}|}$ d'après le medoid le plus proche
- Calculer la nouvelle performance *e.g.* $\tilde{ARI}$ localement au cluster k

$$\tilde{ARI}(k) = \frac{1}{\binom{n_k}{2}} \sum_{(i,j) \in \mathrm{paire}(\mathcal{T})} \mathbb{1}(\tilde{\hat{c}}^{(i)} = \tilde{\hat{c}}^{(j)} = k)\mathbb{1}(y^{(i)} = y^{(j)})$$

- MDARI d'une variable X_j local au cluster k

$$MDARI(X_j, k) = ARI(k) - \tilde{ARI}(k)$$

Tree SHAP [7]

```
ExpectedPred(x,S,Tree)
    procedure G(j, w)
        if v_j ≠ internal then
            return w · v_j
        else
            if d_j ∈ S then
                if x_{d_j} ≤ t_j then
                    return G(a_j, w)
                else
                    return G(b_j, w)
            else
                return
                  G(a_j, w r_{a_j}/r_j) + G(b_j, w r_{b_j}/r_j)
    return G(1, 1)
```

avec

- v_j : valeur feuille j. Si j nœud interne : $v_j =$ "internal"
- a_j, b_j : indexes gauche et droite du nœud j
- t_j : seuil du nœud j
- d_j : index de la caractéristique du nœud j
- r_j : nb. instances dans nœud j

7. Scott M. Lundberg, Gabriel G. Erion et Su-In Lee. “Consistent Individualized Feature Attribution for Tree Ensembles”. (2018).

Similarity Tree SHAP

ExpectedSim($x^{(l)}$, $x^{(m)}$, S, `Tree`)

procedure $P(j, w)$
  if $v_j \neq \textit{internal}$ **then**
    return $w \cdot 1$; `/* the pair reaches a leaf, they are similar */`
  else
    if $d_j \in S$ **then**
      if $(x^{(l)}_{d_j} \leq t_j)$ **and** $(x^{(m)}_{d_j} \leq t_j)$ **then**
        return $P(a_j, w)$
      else if $(x^{(l)}_{d_j} > t_j)$ **and** $(x^{(m)}_{d_j} > t_j)$ **then**
        return $P(b_j, w)$
      else
        return $w \cdot 0$; `/* the node splits the pair, they are dissimilar */`
    else
      return $P(a_j, w\binom{r_{a_j}}{2}/\binom{r_j}{2}) + P(b_j, w\binom{r_{b_j}}{2}/\binom{r_j}{2})$
return $P(1, 1)$