HAL open science

Forêts aléatoires prudentes : une nouvelle stratégie de décision et quelques expériences

Haifei Zhang, Benjamin Quost, Marie-Hélène Masson

▶ **To cite this version:**

Haifei Zhang, Benjamin Quost, Marie-Hélène Masson. Forêts aléatoires prudentes : une nouvelle stratégie de décision et quelques expériences. 31ièmes Rencontres Francophones sur la Logique Floue et ses Applications (LFA 2022), Oct 2022, Toulouse, France. pp.95-101. hal-03895130

HAL Id: hal-03895130

https://hal.science/hal-03895130

Submitted on 9 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Forêts aléatoires prudentes : une nouvelle stratégie de décision et quelques expériences

Haifei Zhang[1] Benjamin Quost[1] Marie-Hélène Masson[1,2]

[1] UMR CNRS 7253 Heudiasyc, Université de Technologie de Compiègne, 60200, France
[2] Université de Picardie Jules Verne, IUT de l'Oise, Beauvais, 60000, France

haifei.zhang@hds.utc.fr benjamin.quost@hds.utc.fr mmasson@hds.utc.fr

Résumé :
Les forêts aléatoires sont une stratégie de classification efficace qui estime les probabilités *a posteriori* des classes en calculant la moyenne des probabilités fournies par les arbres. Lorsque les données d'entraînement sont peu nombreuses, cette estimation devient difficile. Le modèle de Dirichlet imprécis peut être utilisé pour rendre l'estimation robuste en fournissant des intervalles de probabilités en sortie de chaque arbre. Se pose alors la question de l'agrégation des ces intervalles. Nous proposons ici une nouvelle stratégie d'agrégation basée sur la théorie des fonctions de croyance. Nous proposons également d'attribuer des poids aux arbres en fonction de leur degré d'incertitude lors de la classification d'une nouvelle instance. Notre approche est comparée expérimentalement à l'approche de base (moyenne des intervalles) sur plusieurs jeux de données.

Mots-clés :
Classification prudente, forêts aléatoires, modèle de Dirichlet imprécis, fonctions de croyance.

Abstract:
Random forests are an accurate classification strategy which estimates the posterior probabilities of the classes by averaging the outputs provided by the trees. When data are scarce, this estimation becomes difficult. The Imprecise Dirichlet Model can be used to make the estimation robust, providing intervals of probabilities as outputs. Here, we propose a new aggregation strategy based on the theory of belief functions. We also propose to assign weights to the trees according to the amount of uncertainty when classifying a new instance. Our approach is compared experimentally to the baseline approach (average of intervals) on several datasets.

Keywords:
Cautious classification, random forests, imprecise Dirichlet model, belief functions.

1 Introduction

L'apprentissage d'ensemble peut améliorer considérablement les performances des modèles d'apprentissage automatique. Nous pouvons citer trois approches principales : le bagging [5], le stacking [21] et le boosting [13]. Dans une forêt aléatoire [6], une variante du bagging, un grand nombre d'arbres de décision non élagués sont entraînés en introduisant un caractère aléatoire dans l'échantillon ; les sorties de l'arbre (décisions ou distributions de probabilité de classe) sont ensuite agrégées soit par vote, soit par moyennage. Différents schémas d'agrégation ont été comparés dans [16]. Les forêts aléatoires ont été appliquées avec succès dans de nombreux contextes. Cependant, la réalisation de prédictions précises est discutable lorsque l'information disponible est pauvre (par exemple dans les zones de l'espace où il y peu d'exemples d'apprentissage), ou lorsqu'il existe un conflit important entre les sorties des arbres de décision. Dans ce cas, une alternative consiste à garder le modèle prudent en produisant non pas une décision précise mais possiblement un ensemble de décisions. Cette stratégie prudente est liée à la notion de décision avec rejet [7].

Dans cet article, nous traitons le cas où les arbres dans la forêt aléatoire fournissent des prédictions imprécises en utilisant le modèle de Dirichlet imprécis (IDM) de Walley [20]. Alors que l'inférence classique est basée uniquement sur des estimations de probabilités *a posteriori* obtenues en calculant les fréquences de classe sur les instances rattachées à un nœud feuille, l'IDM incorpore une information imprécise sur ces fréquences de classe sous la forme d'un ensemble de distributions de Dirichlet ; par conséquent, par conjugaison, l'infor-

mation *a posteriori* sur les classes est un ensemble actualisé de distributions de Dirichlet [4]. L'information inférée des données dans un nœud feuille peut alors être décrite par des bornes inférieures et supérieures sur les probabilités *a posteriori* des classes. L'intérêt d'utiliser l'IDM dans les arbres de décision afin de fournir des décisions prudentes a été démontré dans plusieurs études [2, 15, 19].

Comme les prédictions générées par une forêt aléatoire prudente sont un ensemble d'intervalles de probabilité, l'agrégation des sorties de l'arbre est plus difficile qu'avec des prédictions précises. Plusieurs stratégies peuvent être utilisées. L'une consiste à obtenir d'abord la décision de chaque arbre (en utilisant par exemple la dominance d'intervalle [18]), puis à obtenir une prédiction finale par vote majoritaire ou vote majoritaire pondéré [1]. Une autre consiste à fusionner directement tous les intervalles de probabilité de l'ensemble, soit en utilisant une disjonction ou une conjonction [8], soit en calculant la moyenne [12], puis à prendre une décision basée sur l'intervalle de probabilité résultant.

Dans cet article, nous proposons une autre approche pour combiner les sorties des arbres, basée sur la théorie des fonctions de croyance [9, 17]. Nous considérons que les arbres fournissent des éléments de preuve sur la valeur des probabilités sous la forme d'intervalles aléatoires fermés définis sur $[0;1]$. Il est alors facile de calculer la croyance et la plausibilité de tout événement défini sur $[0;1]$, et de les utiliser dans un processus de décision prudent [11]. Nous étudions également l'intérêt de pondérer les arbres de l'ensemble, via deux stratégies : des poids basés sur le nombre d'échantillons dans les feuilles, ou des poids égaux à la longueur des intervalles (incertitude épistémique des intervalles). Cette approche peut être vue comme une généralisation du vote pondéré.

2 Préliminaires

2.1 Modèle de Dirichlet imprécis (IDM)

Supposons un espace d'échantillonnage $\Omega = \{\omega_1, \ldots, \omega_K\}$ avec $K \geq 2$ éléments, et soit $\pi_1, \ldots, \pi_K$ un vecteur de probabilités (inconnu) associé, avec $\pi_j = \mathbb{P}(\omega_j)$ pour $j = 1, \ldots, K$. Soit N instances échantillonnées indépendamment de cette distribution : on obtient un vecteur $n = (n_1, \ldots, n_K)$ (avec $\sum_{j=1}^{K} n_j = N$) qui suit une loi multinomiale $\mathscr{M}(n; \pi_1, \ldots, \pi_K)$. En combinant l'*a priori* de Dirichlet imprécis et la vraisemblance des observations, on obtient les probabilités *a posteriori* inférieure et supérieure suivantes :

$$\underline{\mathbb{E}}(\pi_j|n) = \frac{n_j}{N+s}, \quad \overline{\mathbb{E}}(\pi_j|n) = \frac{n_j+s}{N+s}. \quad (1)$$

Le paramètre s peut être interprété comme un nombre d'échantillons virtuels dont l'information sur la classe est inconnue. Bien que plusieurs études aient été menées concernant le choix d'une valeur appropriée [3], ce problème reste encore ouvert.

2.2 Fonctions de croyance induites par des intervalles aléatoires

Soient U et V deux variables aléatoires telles que $U \leq V$ et l'intervalle aléatoire $[U;V]$ associé ; on définit alors une fonction de croyance et de plausibilité sur $\mathbb{R}$ de la manière suivante :

$$\begin{aligned} bel(A) &= \mathbb{P}([U,V] \subseteq A), \\ pl(A) &= \mathbb{P}([U,V] \cap A \neq \emptyset), \end{aligned} \quad (2)$$

pour tout élément A de la σ-algèbre de Borel $\mathscr{B}(\mathbb{R})$ de $\mathbb{R}$ [10]. Soit $I_i = [u_i, v_i]$ avec $i = 1, \ldots, n$, et soit m la fonction de masse de l'ensemble $\mathscr{I}$ des intervalles réels fermés de $[0,1]$ tels que $m(I_i) = m_i$ avec $i = 1, \ldots, n$ et $\sum_{m=1}^{n} m_i = 1$. Dans ce cadre, la croyance et la plausibilité de A sont

$$\begin{aligned} bel(A) &= \sum_{I_i \subseteq A} m_i, \\ pl(A) &= \sum_{I_i \cap A \neq \emptyset} m_i, \quad \forall i = 1, \ldots, n. \end{aligned} \quad (3)$$

Les intervalles I_i tels que $m_i > 0$ sont appelés éléments focaux de m [11]. Cette définition fournit une base pour la mise en commun des éléments d'information fournis par les arbres en ce qui concerne les probabilités des classes.

3 Forêts aléatoires prudentes

Nous nous concentrons ici sur un problème de classification binaire. En supposant un ensemble de données d'apprentissage (x_i, y_i) avec $i = 1, \ldots, N$, où $y_i \in \{0, 1\}$, la probabilité que l'échantillon x_i appartienne à la catégorie 0 (respectivement, 1) s'écrit $p_{i,0}$ (resp., $p_{i,1}$).

Nous considérons une forêt aléatoire composée de T arbres $\{C_1, \ldots, C_t, \ldots, C_T\}$. Dans l'espace des caractéristiques, chaque instance x_i appartient à une certaine région de la forêt aléatoire, définie par l'ensemble des régions $R_i = \{R_i^1, \ldots, R_i^t, \ldots, R_i^T\}$, avec R_i^t la région associée à la feuille dans laquelle l'instance tombe pour l'arbre C_t. L'information sur la feuille est résumée par (n_i^t, N_i^t), où n_i^t est le nombre d'échantillons de catégorie 1 et N_i^t est le nombre total d'échantillons d'apprentissage (l'information étant disponible pour toutes les régions R_i^t, $t = 1, \ldots, T$).

L'IDM donne une estimation sous la forme d'intervalle $I_i^t = [\underline{p}_{i,1}^t, \overline{p}_{i,1}^t]$ de $p_{i,1}$:

$$I_i^t = \left[\frac{n_i^t}{N_i^t + s}, \frac{n_i^t + s}{N_i^t + s}\right] \quad t = 1, \ldots, T. \tag{4}$$

Nous proposons d'agréger ces intervalles en calculant la croyance et la plausibilité que $p_{i,1} \in [0.5; 1]$, c'est-à-dire que les éléments d'information disponibles pointent vers la classe 1 pour l'instance x_i. Selon l'équation (2), nous avons

$$\begin{aligned} bel_{i,1} &= bel(p_{i,1} \in [0.5, 1]) \\ &= \sum_{I_i^t \subseteq [0.5,1]} m_i^t \\ &= \sum_{t=1}^{T} m_i^t \mathbb{1}(\underline{p}_{i,1}^t \geq 0.5), \end{aligned} \tag{5}$$

et

$$\begin{aligned} pl_{i,1} &= pl(p_{i,1} \in]0.5, 1]) \\ &= \sum_{I_i^t \cap]0.5,1] \neq \emptyset} m_i^t \\ &= \sum_{t=1}^{T} m_i^t \mathbb{1}(\overline{p}_{i,1}^t > 0.5), \end{aligned} \tag{6}$$

où m_i^t $(t = 1, \ldots, T)$ est la masse de l'intervalle I_i^t dans le processus d'agrégation et $\mathbb{1}(\cdot)$ est la fonction indicatrice. Notons que, par dualité, nous avons $bel_{i,0} = 1 - pl_{i,1}$ et $pl_{i,0} = 1 - bel_{i,1}$.

Pour la masse, un choix naturel consiste à donner le même poids à chaque arbre, soit $m_i^t = 1/T$; nous proposons ici deux alternatives, qui dépendent du nombre d'instances dans les feuilles. La première est

$$m_i^t = \frac{N_i^t}{\sum_{j=1}^{T} N_i^j}, \quad \forall t = 1, \ldots, T. \tag{7}$$

En utilisant l'équation (7), les feuilles avec moins d'échantillons ont moins de poids dans l'agrégation. La seconde définit, pour $t = 1, \ldots, T$,

$$m_i^t = \frac{1 - u_i^t}{\sum_{j=1}^{T} (1 - u_i^j)}, \text{ avec } u_i^t = \frac{s}{N_i^t + s}, \tag{8}$$

où u_i^t est le niveau d'incertitude épistémique pour l'instance x_i et pour le t-ième arbre. Intuitivement, avec cette proposition, les feuilles ayant une plus petite incertitude épistémique obtiennent des poids plus importants.

Une décision peut alors être prise en appliquant la dominance d'intervalle à $bel_{i,1}$ et $pl_{i,1}$:

$$\hat{y} = \begin{cases} 1 & \text{, si } bel_{i,1} \geq 0.5, \\ 0 & \text{, si } pl_{i,1} < 0.5; \\ \{0, 1\}, & \text{sinon.} \end{cases} \tag{9}$$

4 Expériences

4.1 Modèles de comparaison et critères d'évaluation

Nous avons comparé les performances de notre stratégie avec une approche de référence, qui

consiste à calculer la moyenne des bornes de probabilité inférieure et supérieure sur tous les arbres : $\underline{p}_{i,1} = ave(\underline{p}^t_{i,1})$ et $\overline{p}_{i,1} = ave(\overline{p}^t_{i,1})$. Ce classifieur de base produit également des prédictions indéterminées : la classe 0 est choisie si $\overline{p}_{i,1} < 0,5$ et la classe 1 si $\underline{p}_{i,1} \geq 0,5$, la décision étant indéterminée sinon. Cette approche peut être considérée comme une généralisation de la moyenne des probabilités précises et il a été démontré qu'elle donne de bons résultats [12]. Appliquer un opérateur de combinaison conjonctif ou disjonctif semble peu approprié, les arbres n'étant pas indépendants, et ces opérateurs ne se comportant pas de manière satisfaisante en présence d'intervalles incohérents.

Dans nos expériences, nous avons comparé les deux propositions sur huit ensembles de données provenant du référentiel d'apprentissage machine de l'UCI [14]. Différentes valeurs ont été considérées pour le paramètre IDM : $s \in \{1,3,5\}$, afin d'étudier le comportement des différents schémas d'agrégation par rapport à l'incertitude épistémique.

Nous avons comparé les modèles en utilisant le critère u_{65} [22], qui est couramment utilisé pour comparer des résultats de classification prudente. Dans le cas de la classification binaire, u_{65} récompense les décisions précises et correctes par $\frac{1}{N}$, et les décisions indéterminées par $\frac{0.65}{N}$, ce qui rend la prudence attrayante tout en pénalisant les prédictions indéterminées. La valeur moyenne de u_{65} a été calculée en utilisant la validation croisée en 10 blocs, en raison de la petite quantité de données. Nous avons évalué statistiquement les différences dans les valeurs de u_{65} en répétant cette procédure 50 fois. Tout au long des expériences, la forêt aléatoire était toujours composée de 100 arbres de décision entraînés à la profondeur maximale possible.

4.2 Résultats et discussions

Les tableaux 1, 2 et 3 montrent les valeurs de u_{65} sous forme de pourcentage. Dans ces tableaux, "masse1" correspond à $m^t_i = 1/T$ (tous les arbres ont un poids égal), "masse2" à l'équation (7) (les poids sont proportionnels au nombre d'échantillons dans la feuille), et "masse3" à l'équation (8) (les poids dépendent de l'incertitude épistémique). Nous avons comparé le meilleur résultat obtenu avec l'une ou l'autre des approches crédales à celui de la méthode de base, le meilleur résultat étant indiqué en gras. Nous avons utilisé un test de Student sur les 50 résultats obtenus dans chaque cas, le niveau de signification étant indiqué par des étoiles (1,2,3 étoiles correspondent à des p-valeurs inférieures ou égales à 0,1, 0,05 et 0,01 respectivement).

La stratégie basée sur les fonctions de croyance est légèrement moins performante que la méthode de base pour les faibles valeurs de s. Cependant, lorsque s augmente, ce constat s'inverse. Bien que nous ne puissions pas trouver une stratégie de pondération donnant systématiquement les meilleures performances, les pondérations basées sur l'incertitude épistémique semblent garantir de meilleurs résultats pour les valeurs élevées de s. Rappelons que dans l'IDM, le paramètre s est interprété comme le nombre d'échantillons virtuels avec une catégorie non observée. Par conséquent, des valeurs plus élevées induisent une plus grande incertitude. L'approche fondée sur la théorie des fonctions de croyancee combinée à une stratégie de pondération semble résister à une plus grande incertitude épistémique, en maintenant les performances à un niveau satisfaisant tout en restant déterminées.

Les trois différentes stratégies de pondération peuvent être comparées sur la base de ces résultats préliminaires. Donner des poids égaux aux arbres ("masse1") semble donner sensiblement les mêmes résultats que la méthode de base — notamment en ce qui concerne la robustesse à l'incertitude épistémique. Pondérer les arbres en fonction du nombre d'échantillons dans les feuilles ("masse2") ne semble pas apporter systématiquement de meilleurs résultats

TABLEAU 1 – Comparaison en termes de u_{65} avec s=1

Dataset	masse1	masse2	masse3	baseline
Pima	77.37	76.71	77.30	**78.11****
Heart	82.98	83.17	83.13	**83.56**
Biodeg	87.26	85.49	87.11	**87.67***
Breast-Cancer	**96.19**	95.15	96.07	96.14
Cardiac	77.96	77.98	77.96	**78.12**
Wine	82.42	79.46	82.34	**82.74**
Magic	94.52	93.53	94.41	**94.68**
Spam	95.36	94.33	95.22	**95.50***

TABLEAU 2 – Comparaison en termes de u_{65} avec s=3

Dataset	masse1	masse2	masse3	baseline
Pima	**78.33**	77.32	78.19	78.22
Heart	83.06	83.70	**83.89*****	82.56
Biodeg	**87.99**	85.69	87.64	87.80
Breast-Cancer	96.15	95.10	**96.03**	95.96
Cardiac	78.59	78.80	**78.90**	78.75
Wine	81.86	79.06	**82.07*****	81.45
Magic	**94.32**	93.57	94.28	94.23
Spam	**95.27**	94.32	95.14	95.26

TABLEAU 3 – Comparaison en termes de u_{65} avec s=5

Dataset	masse1	masse2	masse3	baseline
Pima	77.82	77.62	**78.59*****	76.86
Heart	81.17	**83.82*****	83.57	80.47
Biodeg	87.58	85.83	**87.73*****	86.91
Breast-Cancer	95.78	95.08	**95.81**	95.58
Cardiac	78.03	**78.80**	78.44	78.41
Wine	80.57	79.71	**81.53*****	79.62
Magic	93.76	93.52	**94.09*****	93.60
Spam	94.92	94.32	**95.04*****	94.81

que la méthode de base en termes de u_{65}, même pour des valeurs élevées de s. La pondération en fonction du niveau d'incertitude épistémique ("masse3") s'avère beaucoup plus fructueuse dans ce cas, même si ces deux stratégies ("masse2" et "masse3") sont basées sur la même information (nombre d'échantillons dans les feuilles atteint dans chaque arbre par l'instance de test).

5 Conclusion

Dans cette courte contribution, nous avons décrit une nouvelle approche pour agréger les intervalles de probabilité *a posteriori* des classes produites par une forêt aléatoire imprécise. Nous nous sommes concentrés sur les arbres de décision binaires avec des sorties obtenues à l'aide du modèle de Dirichlet imprécis. L'approche est formalisée dans le cadre de la théorie des fonctions de croyance. La croyance que l'instance appartient à la classe positive est estimée comme la proportion d'intervalles soutenant exclusivement l'hypothèse $p_1 \geq 0.5$ (c'est-à-dire avec une borne inférieure dont la valeur est supérieure à 0.5) ; la plausibilité de la classe positive est calculée comme la proportion d'intervalles qui ne contredisent pas cette hypothèse (c'est-à-dire avec une borne supérieure dont la valeur est plus élevée que 0.5). Une décision est ensuite prise en appliquant la dominance d'intervalle. Nous proposons deux variantes en attribuant des poids spécifiques aux arbres, en fonction de la quantité d'information associée à la feuille atteinte par l'échantillon testé. Les variantes de notre stratégie sont comparées à l'approche de base (qui fait la moyenne des intervalles de probabilité) sur plusieurs ensembles de données. En présence d'une faible incertitude, l'approche de base semble être plus performante, tandis que notre approche basée sur des poids dérivés de la quantité d'incertitude semble être beaucoup plus robuste lorsque l'incertitude augmente.

Les travaux futurs peuvent être menés dans plusieurs directions. Premièrement, nous envisagerons d'apprendre automatiquement les poids des arbres en utilisant les données d'entraînement, de manière similaire à l'approche proposée dans [19]. Ensuite, notre objectif est de fournir un moyen d'interpréter les résultats de la forêt aléatoire, et en particulier d'expliquer pourquoi une décision est indéterminée. Enfin, nous étudierons l'extension de notre approche au cas multiclasse.

Déclaration

Cet article est une traduction française de [23].

Références

[1] J. Abellán and A. R. Masegosa. Bagging decision trees on data sets with classification noise. *In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. vol. 5956 LNCS, pp. 248–265, 2010.

[2] J. Abellán and A. R. Masegosa. Imprecise classification with credal decision trees. *Int. J. Uncertainty, Fuzziness Knowlege-Based Syst*. vol. 20, no. 5, pp. 763–787, 2012.

[3] J. Abellán, S. Moral, M. Gómez, and A. R. Masegosa. Varying parameter in classification based on imprecise probabilities. *Advances in Soft Computing*. 37 :231–239, 2006.

[4] J. M. Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*. 39(2-3) :123–150, 2005.

[5] L. Breiman. Bagging predictors. *Machine Learning*. 24(2) :123–140, 1996.

[6] L. Breiman. Random forests. *Machine Learning*. 45(1) :5–32, 2001.

[7] C. Chow. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*. 16(1) :41-46, 1970.

[8] L. De Campos, J. Huete, and S. Moral. Probability interval : A tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 2(2) :167–196, 1994.

[9] A. P. Dempster. Upper and Lower Probabilities Induced by a Multivalued Mapping. *The Annals of Mathematical Statistics*. 38(2) :325–339, 1967.

[10] A. P. Dempster. Upper and Lower Probabilities Generated by a Random Closed Interval. *The Annals of Mathematical Statistics*. 39(3) :957–966, 1968.

[11] T. Denœux. Extending stochastic ordering to belief functions on the real line. *Information Sciences*. 179(9) :1362–1376, 2009.

[12] P. Fink. *Ensemble methods for classification trees under imprecise probabilities*. PhD thesis, Ludwig-Maximilians University, 2012.

[13] Y. Freund and R. E. Schapire. A Decision Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*. 55(1) :119–139, 1997.

[14] M. Lichman. Uci machine learning repository. 2013.

[15] C. J. Mantas and J. Abellán. Credal-C4.5 : Decision tree based on imprecise probabilities to classify noisy data. *Expert Systems with Applications*. 41 : 4625–4637, 2014.

[16] A. J. Sage, U. Genschel, and D. Nettleton. Tree aggregation for random forest class probability estimation. *Statistical Analysis and Data Mining : The ASA Data Science Journal*. 13(2) :134–150, 2020.

[17] G. Shafer. *A mathematical theory of evidence*. Princeton university press. volume 42, 1976.

[18] M. C.M. Troffaes. Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*. 45(1) :17–29, 2007.

[19] L. V. Utkin, M. S. Kovalev, and F. P.A. Coolen. Imprecise weighted extensions of random forests for classification and regression. *Applied Soft Computing Journal*. 92 :106324, 2020.

[20] P. Walley. Inferences from Multinomial Data : Learning About a Bag of Marbles. *Journal of the Royal Statistical Society : Series B (Methodological)*. 58(1) :3–34, 1996.

[21] D. H. Wolpert. Stacked generalization. *Neural Networks*. 5(2) :241–259, 1992.

[22] M. Zaffalon, G. Corani, and D. Mauá. Evaluating credal classifiers by utility-discounted predictive accuracy. *In International Journal of Approximate Reasoning* (53)1282–1301, 2012.

[23] H. Zhang, B. Quost, and MH. Masson. Cautious Random Forests : a new decision strategy and some experiments. *International Symposium on Imprecise Probability : Theories and Applications.* PMLR, p. 369–372, 2021.