



HAL
open science

Modèle crédibiliste pour l'échantillonnage en apprentissage actif

Daniel Zhu, Arnaud Martin, Jean-Christophe Dubois, Yolande Le Gall,
Vincent Lemaire

► To cite this version:

Daniel Zhu, Arnaud Martin, Jean-Christophe Dubois, Yolande Le Gall, Vincent Lemaire. Modèle crédibiliste pour l'échantillonnage en apprentissage actif. Rencontres francophones sur la logique floue et ses applications, Oct 2021, Paris, France. hal-03327140

HAL Id: hal-03327140

<https://hal.science/hal-03327140>

Submitted on 27 Aug 2021

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.

Modèle crédibiliste pour l'échantillonnage en apprentissage actif

Daniel Zhu¹ Arnaud Martin¹ Jean-Christophe Dubois¹ Yolande Le Gall¹ Vincent Lemaire²

¹ Univ Rennes, CNRS, IRISA, DRUID, rue E. Branly, Lannion, prenom.nom@irisa.fr

² Orange Labs, France, vincent.lemaire@orange.com

Résumé :

En apprentissage automatique, la performance d'un modèle supervisé dépend souvent du volume de données étiquetées. Entraîner un modèle sur un grand nombre de données nécessite donc l'étiquetage de beaucoup d'observations et requiert souvent une expertise coûteuse en temps et en argent. Une des solutions consiste alors à externaliser le travail d'étiquetage à une foule de contributeurs dont les qualifications vis-à-vis du domaine étudié seront nécessairement inégales, occasionnant des erreurs d'étiquetage. L'apprentissage actif propose une solution au problème de coût lié aux étiquettes, en impliquant le classificateur dans le choix des données d'entraînement. Ce paradigme peut être combiné à la théorie des fonctions de croyance afin de modéliser l'incertitude dans l'étiquetage des données. Nous proposons dans cet article une méthode pour sélectionner les observations à étiqueter, fondée sur des entropies crédibilistes.

Mots-clés :

Apprentissage actif, fonctions de croyance, entropies, échantillonnage par incertitude

Abstract:

In machine learning, training a classifier on large dataset requires an important amount of labels which is expensive in terms of human resources and money. A possible solution to this problem is to use crowdsourcing in order to label the data. Although, non-expert people do not always have the knowledge to do their work correctly, leading to introducing errors in labeling. Active learning offers a solution to the labeling cost by making the classifier choose the data it wants to label in order to reach good performance with fewer labels. By combining active learning and belief functions, it becomes possible to model the errors and uncertainty in labels. We propose a new sampling method implying belief entropies.

Keywords:

Active learning, Theory of Dempster-Shafer, Belief functions, Entropies, Uncertainty Sampling

1 Introduction

L'utilisation des techniques d'apprentissage automatique requiert parfois de grandes quantités de données d'entraînement afin de parvenir à de bonnes performances de généralisation. Or, s'il est aujourd'hui facile d'accéder à de grandes bases de données, le travail

d'étiquetage nécessaire en amont auprès de personnes expérimentées demeure long et coûteux [3, 14]. L'apprentissage actif propose une solution au problème de coût lié à l'étiquetage en impliquant le classificateur dans le choix des données d'entraînement. Le principe consiste à étiqueter et à apprendre uniquement les observations qui doivent permettre d'améliorer significativement les connaissances du modèle [3, 14]. Il n'est alors plus nécessaire d'étiqueter l'ensemble de la base de données mais seulement une partie d'entre elles. La théorie de Dempster-Shafer — ou théorie des fonctions de croyance — offre un cadre permettant de modéliser l'incertitude dans l'étiquetage des données [15]. En la combinant à l'apprentissage actif, il devient possible de prendre en compte les situations où l'oracle serait incertain des étiquettes qu'il attribue.

Nous proposons dans cet article une méthode d'apprentissage actif permettant de sélectionner les observations à étiqueter avec l'aide d'entropies définies à partir de fonctions de croyance.

Après avoir présenté dans la section 2 les différentes notions que nous utiliserons par la suite, nous exposons notre proposition dans la section 3. La section 4 détaille le protocole expérimental et nos résultats. Enfin la section 5 présente les conclusions et perspectives.

2 Travaux liés

Cette section présente sommairement les notions et les théories utilisées dans cet article. Après une présentation de l'apprentissage actif,

nous rappelons les bases de la théorie des fonctions de croyance.

2.1 Apprentissage actif

L'apprentissage actif est un paradigme de l'apprentissage automatique supervisé dont la particularité réside dans la construction itérative d'un sous-ensemble d'entraînement [3, 14]. Pour cela, le classificateur désigne une observation x et demande à un expert ou à un oracle de lui donner une étiquette correspondante y – c'est ce que nous appelons une requête d'étiquetage ou simplement une requête. Une fois l'étiquette apprise, le classificateur l'ajoute à ses connaissances, s'y ré-entraîne puis formule de nouvelles requêtes en tenant compte de la dernière étiquette.

En sélectionnant de cette façon les observations jugées les plus pertinentes par le classificateur, il n'est plus nécessaire de l'entraîner sur toutes les données d'entraînement, mais seulement sur une partie d'entre elles. Cela réduit par la même occasion les coûts financier et humain de l'apprentissage puisqu'alors, l'expert n'a pas à étiqueter toutes les observations, mais seulement celles demandées par le classificateur.

Il existe plusieurs scénarii d'apprentissage. L'*échantillonnage sélectif* consiste à travailler sur des données issues d'observations réelles, qu'elles soient synthétisées – à l'aide d'un générateur de nombres pseudo-aléatoires suivant une distribution – ou récoltées – à l'aide de mesures. Dans notre contexte, le classificateur évalue toutes les observations et demande l'étiquette de celle qui lui paraît être la plus pertinente pour améliorer ses connaissances – échantillonnage en *pool*. Il s'agit du cas le plus simple et le plus fréquent compte tenu que nous avons aujourd'hui facilement accès à de grosses quantités de données – c'est l'étiquetage qui est coûteux.

Il existe plusieurs techniques pour évaluer la pertinence d'une observation par rapport aux connaissances du classificateur. Les techniques

d'*échantillonnage par incertitude* consistent à sélectionner les observations qui sont les plus difficiles à classer en prenant en compte les connaissances du modèle. Dans cet article, nous nous intéresserons plus particulièrement à l'échantillonnage fondé sur l'entropie de Shannon. En effet, plus celle-ci est élevée et plus l'incertitude est grande. Ainsi, l'observation maximisant l'entropie est la plus difficile à classer et c'est pourquoi elle est envoyée à l'expert :

$$x^* = \arg \max_{x \in \mathcal{U}} \sum_{y \in \Theta} p(y|x) \log \left(\frac{1}{p(y|x)} \right) \quad (1)$$

où \mathcal{U} désigne l'ensemble des observations qui n'ont pas encore été étiquetées et Θ est l'ensemble des classes.

Il a été observé que, parmi les différentes associations d'échantillonnage avec ce classificateur, c'est l'association entre la régression logistique et l'entropie de Shannon qui donne souvent les meilleurs résultats [16]. Nous utilisons la régression logistique lors de nos expérimentations et considérons par conséquent l'entropie de Shannon comme étant la référence idéale avec laquelle comparer nos résultats.

Hüllermeier et al [8] indiquent que l'incertitude peut être de nature épistémique – c'est-à-dire due à un défaut qualitatif de connaissance – ou aléatoire – qui est due au hasard et qui n'est donc pas réductible. Bien que la distinction ne soit pas toujours pertinente en apprentissage automatique, des techniques suggèrent de ne pas la négliger, notamment en apprentissage actif lorsque l'échantillonnage utilisé repose sur l'incertitude. Il convient alors de privilégier l'incertitude épistémique [12].

Nous présentons ci-après une modélisation de l'incertitude épistémique due à l'ignorance de l'oracle à l'aide des fonctions de croyance.

2.2 Théorie de Dempster-Shafer

La théorie de Dempster-Shafer offre un cadre permettant de modéliser l'incertitude et l'imprécision des données à l'aide des fonc-

tions de croyance [5, 15]. Dans ce contexte, l'ensemble Θ est appelé le *cadre de discernement* et regroupe les hypothèses élémentaires du problème – dans notre cas, il s'agit des classes. Une fonction de croyance est alors définie sur 2^Θ – l'ensemble des parties de Θ – et à valeurs dans $[0, 1]$.

Une fonction de masse m vérifie la condition de normalisation :

$$\sum_{X \in 2^\Theta} m(X) = 1 \quad (2)$$

Elle permet notamment de répartir de façon élémentaire la quantité finie de confiance, de crédibilité ou encore de croyance aux différentes hypothèses $X \in 2^\Theta$ du problème.

En outre, si $m(\emptyset) = 0$ alors le monde est *fermé* et notre connaissance du cadre de discernement est exhaustive. Dans le cas contraire, le monde est *ouvert* et il se peut que des hypothèses élémentaires inconnues soient à envisager.

Il est possible de combiner et d'agréger plusieurs fonctions de masse $(m_i)_{i \in \llbracket 1, n \rrbracket}$ issues de n sources d'informations $(S_i)_{i \in \llbracket 1, n \rrbracket}$ différentes, qui auraient chacune réparti leur confiance entre les hypothèses de façon individuelle et indépendante. Il convient pour cela d'utiliser la règle de combinaison conjonctive de Dempster qui permet d'obtenir une fonction de masse m_\oplus à partir d'autres fonctions de masse. Pour toute hypothèse élémentaire $X \in 2^\Theta$:

$$m_\oplus(X) = \sum_{X_1 \cap \dots \cap X_n = X} \prod_{i=1}^n m_i(X_i) \quad (3)$$

Il peut arriver que la validité de l'hypothèse initialement supposée du monde fermé soit mise à mal par cette règle de combinaison. Il suffit alors de normaliser les masses. Pour toute hypothèse élémentaire $X \in 2^\Theta$:

$$m_{\text{Norm}}(X) = \begin{cases} \frac{m_\oplus(X)}{1 - m_\oplus(\emptyset)} & \text{si } X \neq \emptyset, \\ 0 & \text{sinon.} \end{cases} \quad (4)$$

À partir d'une fonction de masse m , il est possible de définir d'autres fonctions de croyance : la crédibilité Cr_m , la plausibilité Pl_m et la communalité Q_m , définies pour tout $Y \in 2^\Theta$ par :

$$\text{Cr}_m(Y) = \sum_{X \subset Y} m(X) \quad (5)$$

$$\text{Pl}_m(Y) = \sum_{X \cap Y \neq \emptyset} m(X) \quad (6)$$

$$\text{Q}_m(Y) = \sum_{X \in 2^\Theta | Y \subset X} m(X) \quad (7)$$

Enfin, il est possible de construire des probabilités BetP_m qualifiées de pignistiques à l'aide des fonctions de masse. Elles sont définies pour tout $Y \in 2^\Theta$ par :

$$\text{BetP}_m(Y) = \sum_{X \in 2^\Theta - \{\emptyset\}} \frac{\text{Card}(X \cap Y)}{\text{Card}(X)} \frac{m(X)}{1 - m(\emptyset)} \quad (8)$$

Cette théorie a été utilisée à de nombreuses reprises en *crowdsourcing* et a su démontrer son efficacité dans la modélisation des imperfections dans les étiquettes [1].

Étant donné que nous utilisons la théorie de Dempster-Shafer en apprentissage actif, nous confondons par la suite le cadre de discernement Θ avec l'ensemble des classes. La notation $p = \text{Card}(\Theta)$ désignera le nombre de classes.

3 Un modèle crédibiliste pour l'échantillonnage

Nous présentons dans cette partie un modèle fondé sur les fonctions de croyance pour décrire une nouvelle méthode d'échantillonnage indépendante du classificateur choisi.

Il existe une méthode d'échantillonnage crédibiliste proposée par Zhang *et al.* [17] qui s'appuie sur les densités et les distances entre les observations. Notre méthode s'en distingue par l'utilisation du modèle d'Appriou défini à partir du classificateur utilisé.

Goëau *et al.* [7] proposent un module de classification interactif faisant intervenir l'apprentissage actif. Les multiples méthodes d'échantillonnage sont toutes fondées sur les fonctions de masse ou indirectement *via* leur probabilité pignistique. Ces dernières sont générées en fonction de la classe des observations voisines [6] ou de leurs attributs.

Ma *et al.* [10] proposent une méthode proche de l'apprentissage actif pour les arbres de décision sur le découpage des nœuds en réduisant l'incertitude épistémique à l'aide d'une variante crédibiliste de la vraisemblance probabiliste.

Notre approche diffère de ces techniques par l'utilisation d'entropies écrites à partir de fonctions de masses générées par le modèle d'Appriou qui ne dépend que du choix du classificateur.

Nous présentons quelque entropies crédibilistes de la littérature, puis le modèle d'Appriou qui permettra de les définir. Enfin, nous proposons une méthode d'échantillonnage crédibiliste.

3.1 Entropies crédibilistes

Il existe de nombreuses entropies fondées les fonctions de croyance [9] qui permettent de mieux modéliser l'incertitude, l'imprécision et l'ignorance que les probabilités.

Les entropies de Höhle H_O , de Smets H_t et de Yager H_y font respectivement appel à la crédibilité, la communalité et à la plausibilité :

$$H_O(m) = - \sum_{X \in 2^\Theta} m(X) \log(\text{Cr}_m(X)) \quad (9)$$

$$H_t(m) = - \sum_{X \in 2^\Theta} \log(Q_m(X)) \quad (10)$$

$$H_y(m) = - \sum_{X \in 2^\Theta} m(X) \log(\text{Pl}_m(X)) \quad (11)$$

Les entropies de Dubois-Prade H_d et de Pal H_p utilisent quant à elles les fonctions de masse et les cardinalités :

$$H_d(m) = \sum_{X \in 2^\Theta} m(X) \log(\text{Card}(X)) \quad (12)$$

$$H_p(m) = \sum_{X \in 2^\Theta} m(X) \log\left(\frac{\text{Card}(X)}{m(X)}\right) \quad (13)$$

Enfin, Nguyen H_n propose une entropie utilisant exclusivement les fonctions de masse :

$$H_n(m) = - \sum_{X \in 2^\Theta} m(X) \log(m(X)) \quad (14)$$

Il convient alors de choisir un modèle capable de générer des fonctions de croyance. Nous présentons dans la suite le modèle d'Appriou.

3.2 Modèle d'Appriou

Notons $M_j(x)$ l'information apportée par le classificateur S sur la classe $\theta_j \in \Theta$ concernant l'observation x pour tout $j \in \llbracket 1, p \rrbracket$. Dans notre cas, ces informations correspondent aux masses affectées sur l'hypothèse θ_j que nous cherchons à modéliser.

Nous introduisons au préalable le modèle des singletons [11], où le classificateur ne peut donner des informations que sur Θ :

$$m_j(\{\theta_j\}) = M_j(x) \quad (15)$$

$$\forall X \in 2^\Theta - \Theta \quad m_j(X) = 0 \quad (16)$$

Le modèle d'Appriou [2] étend ce modèle en introduisant un coefficient d'affaiblissement α_j et repose sur les trois axiomes suivants :

1. Les couples (M_j, α_j) sont des sources d'information distinctes dont les éléments de masse non nulle sont $\{\theta_j\}$, son complémentaire $\{\theta_j^c\}$ et Θ ;
2. Si $M_j = 0$ et si M_j est valide (c'est-à-dire que son affaiblissement α_j vaut 1), alors il est certain que θ_j n'est pas vérifiée;
3. Si la probabilité $p(S|\theta_j)$ représente parfaitement la réalité (*i.e.* $\alpha_j = 1$) et si la probabilité $p(\theta_j)$ est disponible, alors la fonction de masse obtenue par la règle de combinaison de Dempster doit coïncider avec celle obtenue par le théorème de Bayes. La probabilité $p(S|\theta_j)$ correspond à la valeur de M_j dans le modèle des singletons.

Le modèle d'Appriou suivant peut être inféré à partir des trois axiomes. Il consiste à affecter des masses uniquement sur $\{\theta_j\}$, $\{\theta_j^c\}$ et Θ :

$$m_j(X) = \begin{cases} \alpha_j M_j & \text{si } X = \{\theta_j\}, \\ \alpha_j(1 - M_j) & \text{si } X = \{\theta_j^c\}, \\ 1 - \alpha_j & \text{si } X = \Theta, \\ 0 & \text{sinon.} \end{cases} \quad (17)$$

Avec :

$$M_j = \frac{Rp(S|\theta_j)}{1 + Rp(S|\theta_j)} \quad (18)$$

Il convient de noter qu'Appriou avait aussi proposé un autre modèle dont la seule différence réside dans la répartition des masses. Il n'est pas étudié dans le cadre de cet article.

Les probabilités $p(S|\theta_j)$ peuvent être obtenues en sortie de tout classificateur – à condition qu'il soit identique à celui utilisé en apprentissage actif. Ainsi, en amont du modèle Appriou, l'apprentissage du classificateur est réalisé sur une partie P_A de la base de données. Ce classificateur génère alors une prédiction Y_B pour une partie distincte P_B de la base de données, de sorte qu'à l'aide de Y_B et des vraies étiquettes, nous puissions construire C_{Norm} , une matrice de confusion normalisée. Les probabilités $p(S|\theta_j)$ correspondent alors aux coefficients de C_{Norm} .

La fonction de masse obtenue à la fin est issue de la combinaison des fonctions $(m_j)_{j \in [1,p]}$ intermédiaires.

3.3 Méthode d'échantillonnage crédibiliste

La méthode d'échantillonnage proposée est une variante de celle faisant intervenir l'entropie de Shannon Eq. (1). En effet, nous proposons d'utiliser les entropies crédibilistes afin de sélectionner les observations les plus pertinentes. Plus l'entropie sera élevée, plus l'incertitude et l'ignorance sur la donnée seront importantes. La figure 1 illustre cette approche d'échantillonnage crédibiliste.

Ainsi, si $H_{\text{Crédib}}$ est l'une des entropies décrites dans la section 3.1, alors la requête d'étiquetage

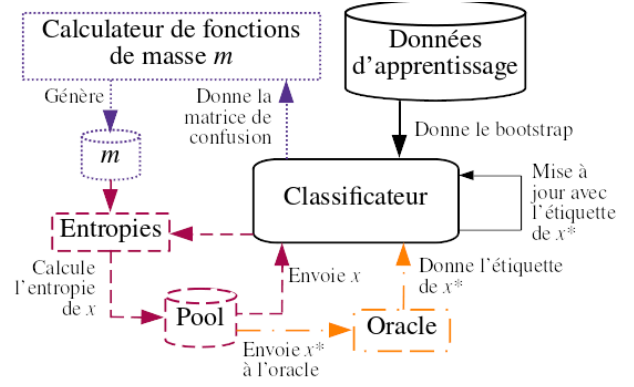


FIGURE 1 – Echantillonnage crédibiliste. Le calcul des fonctions de masse se fait en amont (partie en pointillée). Le classificateur observe l'entropie de toutes les observations x du *pool* (partie en tiret); l'observation x^* maximisant l'entropie est envoyée à l'oracle qui donne l'étiquette au classificateur (partie en tiret-pointillée).

contiendra l'observation qui maximise $H_{\text{Crédib}}$:

$$x^* = \arg \max_{x \in \mathcal{U}} H_{\text{Crédib}}(x) \quad (19)$$

où \mathcal{U} désigne l'ensemble des observations qui n'ont pas encore été étiquetées.

Le calcul des entropies nécessite des fonctions de croyance. Ces fonctions de croyance sont issues du modèle d'Appriou qui peut être appliqué pour tout classificateur.

4 Expérimentations

Nous cherchons à comparer les performances de l'échantillonnage avec les entropies crédibilistes en apprentissage actif en situation de bruit. Le bruit est le phénomène qui remplace la valeur de certaines étiquettes, ce qui les fausse. Cela modélise ainsi l'incertitude de l'oracle lorsqu'il attribue les étiquettes pendant l'apprentissage actif, puisqu'au lieu de donner les vraies étiquettes, il est susceptible d'en attribuer des fausses.

Dans cette section, après avoir introduit et décrit les données et le protocole utilisés pour les expérimentations, nous présentons les résultats

obtenus avec les méthodes d'échantillonnage décrites dans la section 3.

4.1 Données et protocole

Nous nous intéressons à l'apport du modèle d'Appriou lorsque certaines étiquettes sont bruitées. En particulier, nous comparons : l'échantillonnage par l'entropie de Shannon Eq. (1); l'échantillonnage par entropie crédibiliste; l'échantillonnage aléatoire, qui consiste à sélectionner les observations au hasard sans se soucier de leur pertinence pour le classificateur. Cette méthode est régulièrement utilisée pour tester l'efficacité de nouvelles techniques en apprentissage actif.

Les classificateurs utilisés sont la régression logistique (LR), *Random Forest* (RF) ainsi que les 5 plus proches voisins (5NN). Le premier présente une certaine affinité avec l'échantillonnage par entropie [16], le second s'avère particulièrement efficace en apprentissage actif [13] et le dernier est d'usage courant en apprentissage automatique. Nous utilisons le score d'exactitude équilibrée – *balanced accuracy* – afin d'évaluer les performances des différentes méthodes d'échantillonnage [4]. Cette méthode considère mieux les déséquilibres entre les classes que si nous prenions le score d'exactitude classique – *accuracy* – qui peut être biaisé par la classe majoritaire. L'ensemble de test sur lequel le score est calculé contient 30 % des observations.

Nous générons deux bases de données synthétiques contenant 1 000 observations, en dimension 5 et à 2 classes. La première a sa distribution de classes équilibrée et la seconde est distribuée à 80 %-20 %. Nous considérons également trois bases de données issues de l'UCI (voir le tableau 1). Pour chaque base de données, nous étudions différents pourcentages d'étiquettes bruitées (0 %, 10 %, 30 % et 50 %).

Le protocole expérimental est le suivant. Soit $D = \{(x_i, y_i) \mid i \in \llbracket 1, k \rrbracket\}$ une des bases de données introduites précédemment

TABLEAU 1 – Données réelles utilisées dans nos expérimentations. **A** : *Vertebral Column*, **B** : *HCV* et **C** : *Estimation of obesity levels based on eating habits and physical condition*. Les deux dernières lignes correspondent au pourcentage de représentation des classes minoritaire et majoritaire.

Base	A	B	C
Taille	310	615	2 111
Dimension	6	13	16
Classes	2	5	7
% min.	32 %	1,1 %	13 %
% maj.	68 %	87 %	17 %

avec x_i une observation sous la forme d'un vecteur et y_i l'étiquette associée à x_i . Notons $Y = (y_i \mid i \in \llbracket 1, k \rrbracket)$ le k -uplet contenant les étiquettes de chaque observation. Nous commençons par créer Y_b , une copie du k -uplet Y dont b % des étiquettes ont été bruitées ($b \in \{0, 10, 30, 50\}$), puis pour chaque classificateur et chaque méthode d'échantillonnage, nous répétons dix fois le procédé suivant. Nous divisons D en ensemble d'entraînement D_E et de test D_T avec les proportions respectives 70 % et 30 %. Cinq observations de chaque classe sont prélevées de D_E et forment le *bootstrap*, *i.e.* l'ensemble qui pré-entraînera le classificateur. Le restant des données de D_E devient alors le *pool*. Le *bootstrap* est utilisé pour pré-entraîner le classificateur, qui ne connaît donc initialement que cinq observations de chaque classe. Si la méthode d'échantillonnage courante dépend d'entropies crédibilistes, alors nous initialisons le modèle d'Appriou. Dans la phase d'apprentissage actif : tant que l'oracle n'a pas encore satisfait 80 requêtes, le classificateur continue à demander des observations à étiqueter et les confie à l'oracle qui utilise Y_b pour répondre (avec b % des étiquettes fausses). L'observation qui maximise l'entropie courante est la plus incertaine et envoyée à l'oracle. Le classificateur apprend la nouvelle instance et son étiquette puis calcule le score d'exactitude équilibrée, puis réitère.

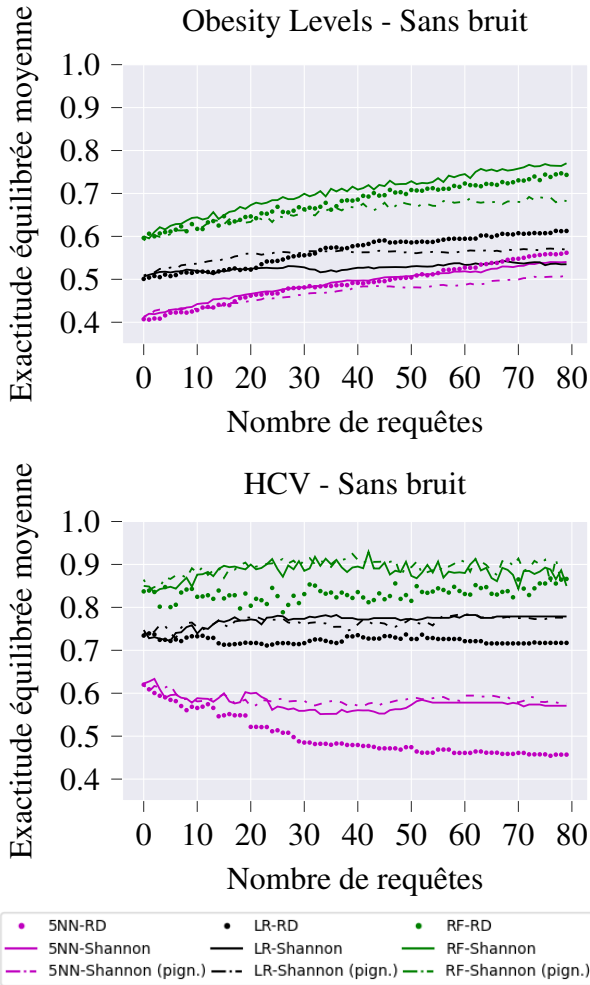


FIGURE 2 – Comparaisons entre les trois classificateurs sur chaque entropie de Shannon (probabilité de classification et pignistique) et l'échantillonnage aléatoire.

4.2 Résultats

Dans un premier temps, la probabilité pignistique Eq. (8) nous permet de comparer deux versions de l'entropie de Shannon : celle dépendant des probabilités de classification et celle dépendant des probabilités pignistiques, obtenues par le modèle d'Appriou. Nous constatons que dans certains cas, la probabilité pignistique peut rivaliser voire surpasser la probabilité de classification en entropie de Shannon ; la figure 2 l'illustre sur quelques bases de données. Le modèle Appriou a donc un apport potentiel sur l'entropie de Shannon ; mais il permet surtout le calcul d'entropies crédibilistes.

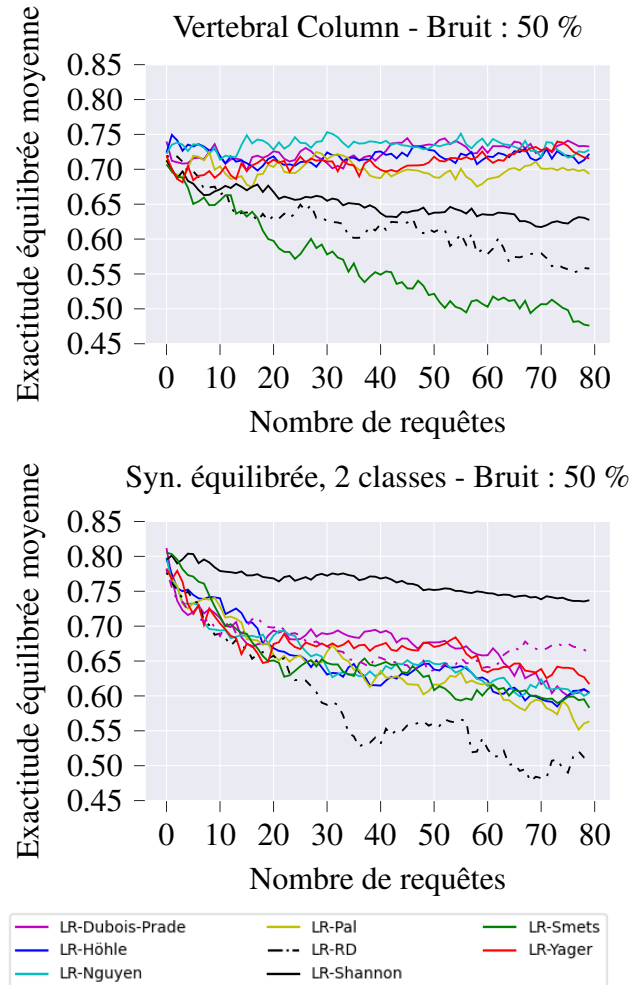


FIGURE 3 – Comparaison entre les différentes entropies crédibilistes sur *Vertebral Column* et les données synthétiques à 2 classes équilibrées.

La figure 3 présente des comparaisons entre entropies crédibilistes avec la régression logistique. Les résultats obtenus ne montrent pas de différence entre les entropies crédibilistes. L'entropie de Shannon reste régulièrement en tête [16]. Cependant nous constatons que les entropies crédibilistes sont plus résistantes face au bruit que l'échantillonnage aléatoire.

5 Conclusion et perspectives

Nous proposons une nouvelle méthode d'échantillonnage en apprentissage actif reposant sur des entropies crédibilistes. Celles-ci sont définies à partir des fonctions de croyance, générées avec le modèle d'Appriou.

Les résultats que nous avons obtenus ne surpassent pas ceux de l'entropies de Shannon. Cependant, il s'avère que les entropies crédibilistes se montrent plus résistantes au bruit sur les étiquettes comparées à une sélection aléatoire. De plus, l'entropie de Shannon peut être alimenté avec les probabilités pignistiques ; ainsi, le modèle d'Appriou offre des perspectives intéressantes en apprentissage actif. Des résultats différents peuvent être obtenus en choisissant un autre modèle génératif de fonctions de masses, tel que celui issu des k plus proches voisins crédibilistes [6].

Références

- [1] L. Abassi and I. Boukhris. A worker clustering-based approach of label aggregation under the belief function theory. *Applied Intelligence*, 49(1) :53–62, 2019.
- [2] A. Appriou. Probabilités et incertitudes en fusion de données multisenseurs. *Revue Scientifique et Technique de la Défense*, 11 :27–40, 1991.
- [3] A. Bondu and V. Lemaire. État de l'art sur les méthodes statistiques d'apprentissage actif. In *AAFD*, 2006.
- [4] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann. The balanced accuracy and its posterior distribution. In *20th International Conference on Pattern Recognition*, pages 3121–3124, 2010.
- [5] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38(2) :325–339, 04 1967.
- [6] T. Denœux. A k -nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(5) :804–813, 1995.
- [7] H. Goëau, M. Rombaut, D. Pellerin, and O. Buisson. Multi-labeled image classification by TBM active learning. In *Belief 2010*, Brest, France, Apr. 2010.
- [8] E. Hüllermeier and W. Waegeman. Aleatoric and epistemic uncertainty in machine learning : an introduction to concepts and methods. *Machine Learning*, 110(3) :457–506, 2021.
- [9] R. Jiroušek and P. P. Shenoy. A new definition of entropy of belief functions in the Dempster–Shafer theory. *International Journal of Approximate Reasoning*, 92 :49–65, 2018.
- [10] L. Ma, S. Destercke, and Y. Wang. Online active learning of decision trees with evidential data. *Pattern Recognition*, 52 :33–45, 2016.
- [11] A. Martin. La fusion d'information. Technical Report Réf. 1484, ENSIETA, 2005.
- [12] V.-L. Nguyen, S. Destercke, and E. Hüllermeier. Epistemic uncertainty sampling. In *Discovery Science*, pages 72–86. Springer International Publishing, 2019.
- [13] C. Salperwyck and V. Lemaire. Learning with few examples : An empirical study on leading classifiers. In *The 2011 International Joint Conference on Neural Networks*, pages 1010–1019, 2011.
- [14] B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin-Madison, 2009.
- [15] G. Shafer. *A mathematical theory of evidence*, volume 42. Princeton university press, 1976.
- [16] Y. Yang and M. Loog. A benchmark and comparison of active learning for logistic regression. *Pattern Recognition*, 83 :401–415, 2018.
- [17] S. Zhang, D. Han, and Y. Yang. Active learning based on belief functions. *Science China Information Sciences*, 63(11) :210205, 2020.