

Apprentissage de rangements prudent avec satisfaction de contraintes

Yonatan Carlos Carranza Alarcón¹

Soundouss Messoudi¹

Sébastien Destercke¹

¹ HEUDIASYC - UMR CNRS 7253, Université de Technologie de Compiègne

57 avenue de Landshut, 60203 COMPIEGNE CEDEX - FRANCE

{yonatan-carlos.carranza-alarcon, soundouss.messoudi, sebastien.destercke}@hds.utc.fr

Résumé :

Apprendre à prédire des rangements d'étiquettes est un problème difficile, en raison de leur nature combinatoire. Une façon de le contourner est de diviser le problème initial en plusieurs sous-problèmes plus simples. Les prédictions obtenues à partir de ces sous-problèmes simplifiés doivent ensuite être combinées en une seule sortie, résolvant les éventuelles incohérences entre les sorties. Dans ce travail, nous adoptons une telle approche en permettant aux sous-problèmes de produire des inférences prudentes sous la forme d'ensembles de rangs lorsque l'incertitude attachée aux données produit des prédictions peu fiables. Plus précisément, nous proposons de combiner une décomposition par rang, dans laquelle chaque sous-problème devient une régression ordinale prudente, avec le problème de satisfaction de contraintes (CSP) pour vérifier la cohérence des prédictions. Nos résultats expérimentaux indiquent que notre approche produit des prédictions avec une fiabilité et une précision équilibrée, tout en restant compétitive avec les approches classiques.

Mots-clés :

Problème de rangement des étiquettes, satisfaction de contraintes, probabilités imprécises.

Abstract:

Ranking problems are difficult to solve due to their combinatorial nature. One way to solve this issue is to adopt a decomposition scheme, splitting the initial difficult problem in many simpler problems. The predictions obtained from these simplified settings must then be combined into one single output, possibly resolving inconsistencies between the outputs. In this paper, we consider such an approach for the label ranking problem, where in addition we allow the predictive model to produce cautious inferences in the form of sets of rankings when it lacks information to produce reliable, precise predictions. More specifically, we propose to combine a rank-wise decomposition, in which every sub-problem becomes an ordinal classification one, with a constraint satisfaction problem (CSP) approach to verify the consistency of the predictions. Our experimental results indicate that our approach produces predictions with appropriately balanced reliability and precision, while remaining competitive with classical, precise approaches.

Keywords:

Label ranking problem, Constraint satisfaction, Imprecise probabilities.

Remerciements :

Ce travail est soutenu par les projets Labex MS2T et PreServe, financé par l'ANR (ANR-11-IDEX-0004-02 et ANR-18-CE23-0008)

1 Introduction

Ces dernières années, les problèmes d'apprentissage automatique avec des sorties structurées ont suscité un intérêt croissant. Ces problèmes apparaissent dans divers domaines, notamment la biologie [23], l'analyse d'images [15], etc.

Dans cet article, nous examinons le *ranking¹ des étiquettes (label ranking, LR)*, où l'on doit associer à chaque instance un ordre complet sur les étiquettes. La plupart des solutions à ce problème réduisent sa complexité initiale, soit en ajustant un modèle probabiliste (Mallows, Plackett-Luce [3]) avec peu de paramètres, ou soit grâce à un schéma de décomposition. Par exemple, le rangement par comparaison par paires (ranking by pairwise comparison RPC) [16] transforme le problème initial en problèmes binaires. La classification des contraintes et les modèles log-linéaires [8], ainsi que les méthodes basées sur les SVM [19] (Support Vector Machine) apprennent, pour chaque étiquette, une fonction d'utilité dont le rang est déduit.

Dans les problèmes de rangement, il peut également être intéressant de prédire des classements partiels plutôt que complets [5, 12], en s'abstenant de faire une prédiction précise lorsqu'on manque d'informations. Ces prédictions peuvent empêcher des décisions néfastes et ont été appliquées dans différents schémas de décomposition, que ce soit par paire [6] ou par étiquette [12], produisant toujours des prédictions prudentes sous la forme de relations

1. Rangement et classement sont utilisés de manière interchangeable pour traduire le terme "ranking".

d'ordre partiel.

Dans cet article, nous proposons une nouvelle méthode de rangement des étiquettes, appelée LR-CSP, basée sur une décomposition par étiquette où chaque sous-problème prédit un ensemble de rangs. Plus précisément, nous proposons d'apprendre pour chaque étiquette un modèle prudent de régression ordinale. Pour cela, nous utilisons le cadre des probabilités imprécises (IP)[20] qui sont bien adaptées pour faire des prédictions prudentes et pour modéliser la manque de connaissances, en décrivant l'incertitude attachée aux données par un ensemble convexe de distributions \mathcal{P} plutôt que par une unique distribution \mathbb{P} .

Un point intéressant de notre méthode, dont le principe peut être utilisé avec n'importe quel ensemble de probabilités, est qu'elle ne nécessite aucune modification du modèle d'apprentissage *imprécis* sous-jacent, tant que ce classifieur peut produire des bornes inférieures et supérieures $[P, \bar{P}]$ sur les problèmes de classification binaire. Nous utilisons ensuite des techniques de CSP sur l'ensemble des prédictions pour vérifier la cohérence de la prédiction du rangement global (c.à.d. que chaque étiquette peut être attribuée à un rang distinct).

La section 2 décrit le cadre général du problème. La section 3 montre comment les rangs peuvent être prédits à partir de modèles probabilistes imprécis et présente la méthode d'inférence proposée. Enfin, la section 4 est consacrée à l'évaluation expérimentale. Des informations supplémentaires se trouvent dans l'article original [2].

2 Description du problème

Les problèmes de classification consistent à associer une instance \mathbf{x} provenant d'un espace d'entrée \mathcal{X} à une seule étiquette de l'espace de sortie $\Lambda = \{\lambda_1, \dots, \lambda_k\}$, représentant les étiquettes possibles. Dans le cas du rangement des étiquettes, une instance \mathbf{x} n'est plus as-

sociée à une étiquette unique de Λ mais à une relation d'ordre² $\succ_{\mathbf{x}}$ sur $\Lambda \times \Lambda$, équivalente à un rangement complet sur les étiquettes dans Λ . Par conséquent, l'espace de sortie est l'ensemble $\mathcal{L}(\Lambda)$ des rangements complets de Λ qui contient $|\mathcal{L}(\Lambda)| = k!$ éléments (c.-à-d. l'ensemble de toutes les permutations). Le tableau 1 illustre un exemple de jeu de données avec $k = 3$ étiquettes.

Tableau 1 – Un exemple d'un jeu de données \mathbb{D} .

X_1	X_2	X_3	X_4	Y
107.1	25	Bleu	60	$\lambda_1 \succ \lambda_3 \succ \lambda_2$
-50	10	Rouge	40	$\lambda_2 \succ \lambda_3 \succ \lambda_1$
200.6	30	Bleu	58	$\lambda_2 \succ \lambda_1 \succ \lambda_3$
107.1	5	Vert	33	$\lambda_1 \succ \lambda_2 \succ \lambda_3$
...

On peut identifier un rangement $\succ_{\mathbf{x}}$ avec une permutation $\sigma_{\mathbf{x}}$ sur $\{1, \dots, k\}$ telle que $\sigma_{\mathbf{x}}(i) < \sigma_{\mathbf{x}}(j)$ ssi $\lambda_i \succ_{\mathbf{x}} \lambda_j$, comme ils sont dans une correspondance biunivoque. $\sigma_{\mathbf{x}}(i)$ est le *rang* de l'étiquette i dans la relation d'ordre $\succ_{\mathbf{x}}$.

Exemple 1 On considère $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ et l'observation $\lambda_3 \succ \lambda_1 \succ \lambda_2$, donc on a $\sigma_{\mathbf{x}}(1) = 2$, $\sigma_{\mathbf{x}}(2) = 3$, $\sigma_{\mathbf{x}}(3) = 1$.

L'objectif usuel du rangement d'étiquette est d'utiliser les données d'entraînement $\mathbb{D} = \{(\mathbf{x}_i, y_i) \mid i=1, \dots, n\}$ avec $x_i \in \mathcal{X}$, $y_i \in \mathcal{L}(\Lambda)$ pour apprendre un prédicteur, ou un *ordonnanceur* $h : \mathcal{X} \rightarrow \mathcal{L}(\Lambda)$. Alors qu'en théorie ce problème peut être transformé en un problème de classification où chaque rangement est une étiquette différente, cela est en pratique impossible, car le nombre d'étiquettes distinctes augmenterait de manière factorielle avec k . Le moyen le plus courant pour résoudre ce problème consiste soit à décomposer le problème en plusieurs sous-problèmes plus simples, ou soit à ajuster une distribution de probabilité paramétrique sur les rangs [3]. Dans cet article, nous nous concentrerons sur une décomposition par étiquette.

2. Une relation complète, transitive et asymétrique.

3 Décomposition par étiquette : apprentissage et prédiction

Nous proposons de réduire le problème de rangement initial en autant de sous-problèmes que d'étiquettes dans Λ , où le rang correspondant à cette étiquette dans l'exemple devient la nouvelle classe du sous-problème (Fig. 1). L'idée est donc d'apprendre un modèle probabiliste $p_i : K \rightarrow [0, 1]$ avec $K = \{1, 2, \dots, k\}$, où $p_{ij} := p_i(j)$ correspond à la probabilité $P(\sigma(i) = j)$ que l'étiquette λ_i ait le rang j , et $\sum_j p_{ij} = 1$.

Une première étape consiste à décomposer les données d'origine \mathbb{D} en k ensemble de données $\mathbb{D}_j = \{(\mathbf{x}_i, \sigma_{\mathbf{x}_i}(j)) \mid i = 1, \dots, n\}$, $j = 1, \dots, k$. L'estimation des probabilités p_{ij} pour une étiquette λ_i revient alors à résoudre un problème de régression ordinale [14]. Dans de tels problèmes, le rang associé à une étiquette est celui qui minimise la perte attendue \mathbb{E}_{ij} pour affecter l'étiquette λ_i au rang j , cela dépend de p_{ij} et d'une distance $D : K \times K \in \mathbb{R}$ entre les rangs comme suit :

$$\mathbb{E}_{ij} = \sum_{\ell=1}^k D(j, \ell) p_{i\ell}. \quad (1)$$

Les choix courants pour les distances sont les normes L_1 et L_2 , correspondant à

$$D_1(j, \ell) = |j - \ell| \quad \text{et} \quad D_2(j, \ell) = (j - \ell)^2.$$

L'intérêt de ces distances et d'autres dans le cadre imprécis, est de produire des prédictions sous forme d'intervalles [11], c.-à-d. que $\{1, 3\}$ ne peut pas être une prédiction mais $\{1, 2, 3\}$ peut l'être. Dans la suite, nous nous concentrons sur la distance L_1 qui est la plus souvent considérée dans la régression ordinale³.

3.1 Modèle d'ensemble de probabilités

Les estimations précises p_i obtenues à partir des données d'apprentissage \mathbb{D}_k peuvent être peu fiables, surtout si ces estimations reposent

3. L'approche s'adapte facilement aux autres distances.

sur des données bruitées ou en petit nombre. Nous proposons plutôt d'utiliser les probabilités imprécises [21] qui modélisent l'incertitude en estimant un ensemble convexe de probabilités (un polytope) pour chaque étiquette λ_i . Dans notre contexte, un modèle imprécis particulièrement intéressant sont les p-boîtes [10], qui encodent la nature ordinale des classements et sont un choix courant dans le cadre précis [14]. Elles consistent à fournir des bornes de probabilités $[\underline{P}(A_\ell), \overline{P}(A_\ell)]$ sur les événements $A_\ell = \{1, \dots, \ell\}$ et à considérer l'ensemble

$$\mathcal{P}_i = \left\{ p_i : \underline{P}_i(A_\ell) \leq \sum_{j=1}^{\ell} p_{ij} \leq \overline{P}_i(A_\ell), \sum_{j \in K} p_{ij} = 1 \right\}. \quad (2)$$

Nous noterons par $\underline{F}_{ij} = \underline{P}_i(A_j)$ et $\overline{F}_{ij} = \overline{P}_i(A_j)$ les bornes de A_j . Dans le tableau 2, nous donnons un exemple de la fonction de répartition (FR) imprécise pour λ_i avec $k = 5$.

Tableau 2 – FR imprécise de λ_i

Rang j	1	2	3	4	5
\overline{F}_{ij}	0.15	0.55	0.70	0.95	1
\underline{F}_{ij}	0.10	0.30	0.45	0.80	1

Cette approche requiert une modélisation de k différents modèles, un pour chaque étiquette, contrairement à l'approche RPC [16] qui en requiert $k(k-1)/2$ (un pour chaque paire d'étiquettes), ce qui est calculatoirement avantageux.

3.2 Inférence par rang

Dans une approche classique, déterminer le rang optimal d'une étiquette revient à comparer les coûts espérés de chaque rang, et à prendre le rang ayant le coût espéré minimum. Cela revient à prendre l'élément maximal d'un ordre \succ sur les rangs tel que $\ell \succ m$, si et seulement si

$$\sum_{j=1}^k D_1(j, m) p_{ij} \geq \sum_{j=1}^k D_1(j, \ell) p_{ij}. \quad (3)$$

Lorsque les probabilités précises p_i sont remplacées par des ensembles de probabilités \mathcal{P}_i , l'approche classique peut être étendue en utili-

$\hat{R}_1^* = \{1\}$ qui se propage ensuite aux autres, etc.

Ce genre de problème est bien connu de la programmation par contraintes [7], où il correspond à la contrainte «toutes différentes».

4 Expériences

Dans cette section, nous effectuons des expériences sur 14 jeux de données extraits de dépôt UCI [13] et Statlog [17] (c.f. Tableau 3), en utilisant 10×10 -fold validation croisée. Les jeux de données d'origine ont été transformés à des jeux de données de classement complet en suivant la procédure décrite dans [4].

Tableau 3 – Jeux de données d'expérimentation

#	Data set	Type	#Inst	#Attributes	#Labels
<i>a</i>	authorship	classification	841	70	4
<i>b</i>	bodyfat	regression	252	7	7
<i>c</i>	calhousing	regression	20640	4	4
<i>d</i>	cpu-small	regression	8192	6	5
<i>e</i>	fried	regression	40768	9	5
<i>f</i>	glass	classification	214	9	6
<i>g</i>	housing	regression	506	6	6
<i>h</i>	iris	classification	150	4	3
<i>i</i>	pendigits	classification	10992	16	10
<i>j</i>	segment	classification	2310	18	7
<i>k</i>	stock	regression	950	5	5
<i>l</i>	vehicle	classification	846	18	4
<i>m</i>	vowel	classification	528	10	11
<i>n</i>	wine	classification	178	13	3

4.1 Exactitude/complétude trade-off

Afin d'identifier correctement sur quelles étiquettes il est souhaitable de s'abstenir ou de fournir un ensemble de rangs possibles, il faut mesurer le degré de précision et d'exactitude des prédictions. En effet, un bon équilibre doit être recherché entre l'informativité et la fiabilité des prédictions. Pour cette raison, de même que dans [5], nous utilisons une mesure d'exactitude (correctness (CR)) et de complétude (completeness CP) pour évaluer la qualité des prédictions. Étant donné la prédiction globale $\hat{R} = \{\hat{R}_i, i = 1, \dots, k\}$;

$$\begin{aligned}
 CP(\hat{R}) &= \frac{k^2 - \sum_{i=1}^k |\hat{R}_i|}{k^2 - k}, \\
 CR(\hat{R}) &= 1 - \frac{\sum_{i=1}^k \min_{\hat{r}_i \in \hat{R}_i} |\hat{r}_i - r_i|}{0.5k^2}
 \end{aligned} \tag{6}$$

où CP est nul si tous les \hat{R}_i contiennent les k rangs possibles et a une valeur de 1 si tous les \hat{R}_i sont réduites en singletons, tandis que CR est équivalent à la Spearman Footrule en cas de prédiction précise.

4.2 Notre approche

Comme indiqué plus tôt, notre proposition consiste à ajuster un modèle de régression ordinaire imprécise pour chaque étiquette dont il doit estimer les bornes inférieures et supérieures de la fonction de répartition $[\underline{F}_i, \overline{F}_i]$. À cet égard, nous proposons d'utiliser une extension de la méthode de Frank et Hall [14] pour des probabilités imprécises. Le méthode de Frank et Hall tire profit de la sortie structurée y (c.-à-d. l'ordonnancement complète des étiquettes) en transformant le problème de régression ordinaire de k étiquettes en $k - 1$ sous-problèmes de classification binaire. Chaque estimation de la probabilité $P_i(A_\ell) := F_i(\ell)$ où $A_\ell = \{1, \dots, \ell\} \subseteq K$ et l'application $F_i : K \rightarrow [0, 1]$ peut être considérée comme une fonction de répartition discrète. Nous proposons simplement de rendre ces estimations imprécises et d'utiliser les bornes suivantes

$$\underline{P}_i(A_j) := \underline{F}_i(j) \quad \text{et} \quad \overline{P}_i(A_j) := \overline{F}_i(j)$$

qui est en effet le modèle de p-box généralisé [10], tel que défini dans l'équation (2).

Pour estimer ces bornes, nous proposons d'utiliser le classifieur naïf credal (NCC)[24], qui étend le classifieur naïf classique de Bayésienne (NBC). Le niveau d'imprécision du classifieur est contrôlé par un hyper-paramètre $s \in \mathbb{R}$. En effet, plus s est élevé, plus les intervalles sont larges $[\underline{P}_i(A_j), \overline{P}_i(A_j)]$. Pour $s = 0$, nous récupérons la NBC classique avec des prédictions précises, et pour $s \gg \gg 0$, le modèle NCC fera des prédictions vides (c.-à-d. tous les rangs pour chaque étiquette)

La valeur d'imprécision s peut être différente d'un jeu de données à l'autre (voir la figure 2), et il est essentiel de pouvoir obtenir deux valeurs extrêmes

- s_{min} : correspondant à la valeur avec une *complétude* moyenne proche de 1, rendant le classifieur proche d'un précis. Cette valeur est celle que nous utiliserons pour comparer notre approche à des approches précises,
- s_{max} : correspondant à la valeur avec une *exactitude* moyenne proche de 1, et pour laquelle les prédictions faites sont presque toujours correctes.

Pour trouver ces valeurs; nous partons d'un intervalle initial de valeurs $[\underline{s}, \bar{s}]$, ainsi que ces intervalles cibles $[\underline{CP}, \overline{CP}]$ et $[\underline{CR}, \overline{CR}]$, généralement $[0.95, 1]$. Notez qu'en cas de prédictions incohérentes, $\hat{R}_i = \emptyset$ et la complétude est supérieure à 1 (on considère $CR=0$). Pour s_{min} , nous partons généralement de $\underline{s} = 0$ ($CP > 1$) et considérons une valeur \bar{s} suffisamment grande pour laquelle $CP < 0.95$ (e.g. en partant de $s = 2$ comme l'a conseillé [22] et en doublant s de manière itérative jusqu'à $CP < 0.95$, car lorsque s augmente, la complétude diminue et l'exactitude augmente en moyenne). On procède ensuite par dichotomie pour trouver une valeur s_{min} pour laquelle les prédictions moyennes sont dans l'intervalle $[\underline{CP}, \overline{CP}]$. Nous procédons de même pour s_{max} .

Avec s_{min} et s_{max} retrouvés, nous procédons à calculer les valeurs intermédiaires (voir la figure 2) via une procédure simple; nous commençons par calculer la CP et la CR de la valeur du milieu de l'intervalle $[s_{min}, s_{max}]$, c.-à-d. $(s_{min} + s_{max})/2$. Nous calculons ensuite la distance entre tous les couples des valeurs de la complétude/exactitude obtenues pour les valeurs s consécutives, et ajoutons un nouveau point s au milieu entre les deux points avec la plus grande distance euclidienne. Nous répétons ce processus autant de fois que le nombre de valeurs s que nous voulons.

Dans la figure 2, nous constatons que l'imprécision liée à la valeur de s dépend de manière significative du jeu de données. Notre approche nous permet de trouver la valeur «optimale» de s_{min} pour chaque jeu

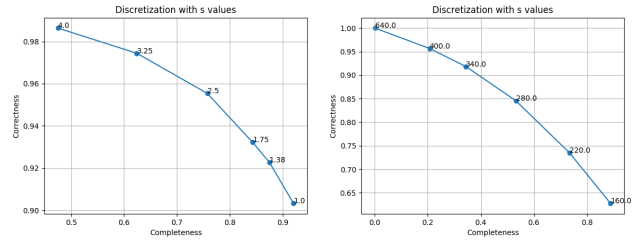


Figure 2 – Évolution de l'hyper-paramètre s sur les jeu de données : Glass (à gauche) et Calhousing (à droite).

de donnée, qui peut être petite ($s_{min} = 1$ pour Glass) ou grande ($s_{min} = 160$ pour Calhousing). Nous constatons aussi que pour le jeu de données Glass on peut avoir une exactitude presque parfaite tout en n'étant pas totalement vide $\hat{R}_i = \emptyset$ (car l'exactitude de presque 1 est atteinte pour une complétude légèrement inférieure à 0.5, quand $s = 4$), alors que cela ne peut pas être le cas pour d'autres jeu de données plus difficiles tels que le Calhousing.

4.3 Comparaison avec d'autres méthodes

Reste à savoir si notre approche est compétitive par rapport aux autres existantes. Pour y parvenir, nous comparons le résultats obtenus de notre approche paramétré à $s = s_{min}$ avec les autres. Les méthodes auxquelles nous nous comparons sont les suivantes :

- Le rangement par comparaison par paires (RPC), mis en oeuvre dans [1];
- L'arbre de rangement des étiquettes (the Label ranking tree (LRT))[4]), qui adopte un schéma local non décomposé.
- Le rangement par étiquettes en utilisant machine à vecteurs de support (SVM-LR), décrit dans [19]⁴.

Vu que NCC manipule des attributs prenant des valeurs discrètes, nous devons donc discretiser les attributs continus à z intervalles avant de procéder à l'apprentissage⁵. Bien que z puisse être optimiser, nous le fixons arbitrairement à $z=6$ pour le comparer aux autres.

4. Disponible en <https://pypi.org/project/svm-label-ranking/>
5. Disponible en <https://github.com/sdestercke/classifip>.

Un dernier problème à résoudre est de savoir comment gérer les prédictions incohérentes, celles dans lesquelles la contrainte «*toutes différentes*» ne trouverait pas une solution précise ou partielle mais une solution vide. Ici, de telles prédictions sont ignorées, et nos résultats ne considèrent que la CR et la Spearman footrule sur des solutions cohérentes, le traitement des prédictions incohérentes fera l’objet de travaux futurs.

4.4 Résultats

Les résultats moyens et leur classement entre parenthèses obtenus en termes de la mesure d’exactitude (CR) sont présentés dans le tableau 4 avec une discrétisation $z=6$ paramétrée à notre approche (appelé LR-CSP-6).

Tableau 4 – Les résultats moyens de CR(%)

	LR-CSP-6	LRT	RPC	SVM-LR
<i>a</i>	93.90 ± 0.69 (1)	91.53 ± 0.31 (3)	93.21 ± 0.23 (2)	64.42 ± 0.36 (4)
<i>b</i>	54.12 ± 3.73 (1)	41.70 ± 1.48 (4)	50.43 ± 0.39 (3)	51.10 ± 0.49 (2)
<i>c</i>	61.05 ± 0.80 (1)	58.37 ± 0.28 (2)	51.85 ± 0.02 (3)	38.45 ± 0.02 (4)
<i>d</i>	68.72 ± 1.42 (1)	60.76 ± 0.30 (3)	61.93 ± 0.04 (2)	46.71 ± 0.87 (4)
<i>e</i>	99.20 ± 0.07 (2)	91.26 ± 0.06 (3)	99.92 ± 0.01 (1)	84.18 ± 2.67 (4)
<i>f</i>	91.95 ± 2.90 (1)	91.59 ± 0.47 (2)	90.83 ± 0.24 (3)	85.68 ± 0.33 (4)
<i>g</i>	79.21 ± 3.37 (2)	85.09 ± 0.46 (1)	74.86 ± 0.16 (3)	70.16 ± 0.46 (4)
<i>h</i>	99.36 ± 1.28 (1)	97.16 ± 0.55 (2)	92.75 ± 0.58 (3)	87.39 ± 0.37 (4)
<i>i</i>	91.31 ± 0.14 (3)	95.14 ± 0.05 (1)	94.12 ± 0.01 (2)	58.75 ± 2.71 (4)
<i>j</i>	91.20 ± 0.85 (3)	96.11 ± 0.10 (1)	94.52 ± 0.03 (2)	66.25 ± 3.05 (4)
<i>k</i>	88.63 ± 1.53 (2)	91.64 ± 0.27 (1)	82.23 ± 0.08 (3)	75.20 ± 0.17 (4)
<i>l</i>	85.29 ± 1.91 (3)	88.03 ± 0.44 (2)	89.24 ± 0.14 (1)	81.93 ± 1.00 (4)
<i>m</i>	88.23 ± 1.00 (1)	84.40 ± 0.62 (2)	72.88 ± 0.06 (3)	65.41 ± 1.21 (4)
<i>n</i>	98.20 ± 1.19 (1)	91.80 ± 0.87 (4)	94.58 ± 0.61 (2)	94.56 ± 0.50 (3)
avg.	85.03 ± 1.49(1.64)	83.18 ± 0.45(2.21)	81.67 ± 0.19(2.36)	69.30 ± 1.02(3.79)

Un test de Friedman [9] sur les rangs donne un p-valeur de 0.0001097, suggérant une différence significative entre les algorithmes. Le test post-hoc de Nemenyi indique en outre que LR-CSP-6 est significativement meilleur que SVM-LR (avec un p-valeur de 0.000066). Notre approche reste également compétitive avec LRT (avec un p-valeur de 0.6450) et RPC (avec un p-valeur de 0.4600).

Nous rappelons enfin que notre méthode est également assez rapide à calculer, grâce à l’utilisation simultanée de la procédure de décomposition, du calcul d’intervalles de probabilités, et de la fonction de perte qui rendent l’étape de prédiction très efficace. De plus, comme nos prédictions sont des intervalles, c.-à-d. des ensembles de rangs sans trous,

nous pouvons utiliser des algorithmes très efficaces pour traiter les contraintes «*toutes différentes*» [18].

5 Conclusion et perspectives

Cet article propose une méthode (LR-CSP) pour faire de prédictions partielles dans le classement des étiquettes, en utilisant une décomposition par étiquette ainsi que des prédictions partielles en termes de rangs possibles. Les expériences sur les jeux de données montrent que notre approche produit des prédictions fiables et prudentes, avec une performance proche et parfois meilleure que les autres. Plusieurs pistes de recherche s’offrent à nous : (1) l’optimisation du paramètre de discrétisation z , (2) un classifieur imprécis de base différent (3) l’appliquer aux classements par paires, etc. Certains de nos futurs travaux porteront sur cela.

Références

- [1] Alexander Balz and Robin Senge. Weka-*lr* : A label ranking extension for weka. 2011.
- [2] Yonatan-Carlos Carranza-Alarcon, Soudouss Messoudi, and Sébastien Des-tercke. Cautious label-wise ranking with constraint satisfaction. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 96–111, 2020.
- [3] W. Cheng, K. Dembczynski, and E. Hüllermeier. Label ranking methods based on the plackett-luce model. In *Proceedings of the 27th Annual International Conference on Machine Learning - ICML*, pages 215–222, 2010.
- [4] W. Cheng, J. Hühn, and E. Hüllermeier. Decision tree and instance-based learning for label ranking. In *Proceedings of the 26th Annual International Conference on Machine Learning - ICML’ 09*, 2009.
- [5] W. Cheng, M. Rademaker, B. De Baets, and E. Hüllermeier. Predicting partial

- orders : ranking with abstention. *Machine Learning and Knowledge Discovery in Databases*, pages 215–230, 2010.
- [6] Weiwei Cheng, Eyke Hüllermeier, Willem Waegeman, and Volkmar Welker. Label ranking with partial abstention based on thresholded probabilistic models. In *Advances in neural information processing systems*, pages 2501–2509, 2012.
- [7] Rina Dechter. *Constraint processing*. Morgan Kaufmann, 2003.
- [8] O. Dekel, C. D. Manning, and Y. Singer. Log-linear models for label ranking. In *Advances in Neural Information Processing Systems*, 2003.
- [9] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan) :1–30, 2006.
- [10] S Destercke, D Dubois, and E Chojnacki. Unifying practical uncertainty representations : I. generalized p-boxes. *International Journal of Approximate Reasoning*, 49 :649–663, 2008.
- [11] Sébastien Destercke. On the median in imprecise ordinal problems. *Annals of Operations Research*, 256(2) :375–392, 2017.
- [12] Sébastien Destercke, Marie-Hélène Masson, and Michael Poss. Cautious label ranking with label-wise decomposition. *European Journal of Operational Research*, 246(3) :927–935, 2015.
- [13] A. Frank and A. Asuncion. UCI machine learning repository, 2010.
- [14] Eibe Frank and Mark Hall. A simple approach to ordinal classification. In *European Conference on Machine Learning*, pages 145–156. Springer, 2001.
- [15] X. Geng. Multilabel ranking with inconsistent rankers. In *Proceedings of CVPR 2014*, 2014.
- [16] E. Hüllermeier, J. Furnkranz, W. Cheng, and K. Brinker. Label ranking by learning pairwise preferences. *Artificial Intelligence*, 172 :1897–1916, 2008.
- [17] R.D. King, C. Feng, and A. Sutherland. Statlog : Comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence*, 9(3) :289–333, 1995.
- [18] Alejandro López-Ortiz, Claude-Guy Quimper, John Tromp, and Peter Van Beek. A fast and simple algorithm for bounds consistency of the alldifferent constraint. In *IJCAI*, volume 3, pages 245–250, 2003.
- [19] Shankar Vembu and Thomas Gärtner. Label ranking algorithms : A survey. In *Preference learning*, pages 45–64. Springer, 2010.
- [20] P. Walley. *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, New York, 1991.
- [21] Peter Walley. The elicitation and aggregation of beliefs. Technical report, University of Warwick, 1982.
- [22] Peter 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.
- [23] N. Weskamp, E. Hüllermeier, D. Kuhn, and G. Klebe. Multiple graph alignment for the structural analysis of protein active sites. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 4(2) :310–320, 2007.
- [24] Marco Zaffalon. The naive credal classifier. *Journal of statistical planning and inference*, 105(1) :5–21, 2002.