

Ordonnement d'alternatives dans le cadre de la théorie des fonctions de croyance

Label ranking using belief functions

M.-H. Masson¹

Li Qiang²

T. Dencœur²

¹ Université de Picardie Jules Verne

² Université de Technologie de Compiègne

Laboratoire Heudiasyc, UMR CNRS 6599

BP 20529 60205 Compiègne

mmasson@hds.utc.fr

Résumé :

L'ordonnement d'alternatives est un problème d'apprentissage dont le but est de relier des instances à un ordre total défini sur un ensemble d'alternatives possibles. Nous abordons ce problème dans le cadre de la théorie des fonctions de croyance. Une approche de décomposition en paires de préférences sur les alternatives est adoptée. Pour chaque paire d'alternatives, on construit un classifieur binaire évidentiel. Ensuite, à partir des informations fournies par les classifieurs binaires, deux méthodes, l'une de vote, l'autre basée sur le maximum de plausibilité d'une permutation, permettent d'obtenir un ordre total des alternatives pour chaque instance. Ces deux méthodes sont comparées sur plusieurs jeux de données.

Mots-clés :

Ordonnement d'alternatives, fonctions de croyance, préférences par paires, classifieur évidentiel.

Abstract:

Label ranking is a learning task where the goal is to map instances to a linear order on a finite set of predefined labels. This problem is tackled in the framework of belief functions theory. A pairwise preference decomposition is used. For each pair of labels, an evidential binary classifier is built. Then, using informations provided by the classifiers, two methods for deriving a total order on the labels for each instance are proposed. The first is a simple voting scheme. The second one looks for the permutation of the labels with maximum plausibility. These methods are compared using various data sets.

Keywords:

label ranking, belief functions theory, pairwise preferences, evidential classifier.

1 Introduction

Le problème de l'ordonnement d'alternatives suscite un intérêt croissant dans la communauté d'apprentissage statistique. Il s'agit d'apprendre, à partir d'exemples, une application

associant des instances x décrites par des caractéristiques dans un espace d'entrée \mathbb{R}^p à un ordre total sur un espace fini de labels (ou d'alternatives) $\mathcal{L} = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$. Par exemple, à partir du profil d'un consommateur, on cherche à prédire ses préférences en termes de marques de voiture (comme illustré Figure 1), de destination de vacances, etc... Le problème de l'ordonnement d'alternatives peut être vu comme une extension du problème conventionnel de discrimination où l'on ne cherche à prédire que le label de premier rang. Ce problème entretient des liens avec d'autres champs de l'apprentissage statistique comme la classification multilabels [7, 15], la régression ordinale [11, 4] ou encore l'ordonnement d'instances (où il s'agit cette fois d'ordonner les instances elles-mêmes) [2, 3, 8]. Les domaines d'application potentiels de l'ordonnement d'alternatives sont nombreux, par exemple la catégorisation de textes, ou encore les systèmes de recommandation.

Beaucoup d'approches permettant l'ordonnement d'alternatives ont été proposées récemment dans la littérature. La *classification sous contraintes* [10] consiste à apprendre des fonctions de score associées à chaque alternative de telle sorte que l'ordre observé sur les scores respecte autant que possible l'ordre sur les alternatives. Les fonctions scores sont des fonctions linéaires ou à noyaux. Elles peuvent être déterminées par des approches du type

Tableau 1 – Le problème de l’ordonnement d’alternatives

Prénom	Age	Profession	Résidence	Préférences
Paul	46	Chercheur	Pierrefonds	Volvo \succ BMW \succ Citroën
Harry	49	Ingénieur	Compiègne	BMW \succ Citroën \succ Volvo
Mylène	23	MdC	Beauvais	Citroën \succ BMW \succ Volvo
Paulette	67	Retraitée	Trosly	?

séparateurs à vaste marge. Dans l’*approche à base de cas* [1], l’idée est de déduire un ordre total à partir des ordres totaux associés aux voisins de l’instance dans l’espace d’entrée grâce à une fonction d’agrégation. Nous nous sommes inspirés d’une autre approche, dite de *préférence par paires* [9, 12]. Elle préconise d’apprendre, pour chaque couple d’alternatives (i, j) , un modèle binaire \mathcal{M}_{ij} permettant de prédire si λ_i est préféré ou non à λ_j . Elle conduit à la résolution de $m(m-1)/2$ problème de discrimination binaires. Un ordre total sur les alternatives s’obtient ensuite par un système assimilable à un vote.

Dans cet article, nous adoptons cette approche par paires en la plaçant dans le cadre théorique des fonctions de croyance. Dans ce qui suit, nous commençons par quelques brefs rappels sur la théorie des fonctions de croyance. Ensuite, après avoir évoqué la méthode de préférence par paires de Hüllermeier, nous présentons la modélisation proposée. Le paragraphe 5 expose quelques résultats expérimentaux. Enfin le paragraphe 6 conclut cet article.

2 Quelques éléments de la théorie des fonctions de croyance

La théorie de l’évidence de Dempster Shafer (ou théorie des fonctions de croyance) [13], est un cadre alternatif à celui des probabilités pour raisonner avec des informations partielles et incertaines. Nous donnons dans ce paragraphe les principaux concepts utiles à la compréhension de l’article.

On considère une variable ω prenant ses valeurs

dans un ensemble fini Ω appelé cadre de discernement. La connaissance partielle quant à la valeur prise par ω peut être représentée par une fonction de masse, [13, 14], définie comme une fonction m de 2^Ω dans $[0, 1]$, vérifiant :

$$\sum_{A \subseteq \Omega} m(A) = 1. \quad (1)$$

Les sous-ensembles A de Ω tels que $m(A) > 0$ sont appelés des *éléments focaux* de m . Chaque élément focal A représente un ensemble possible de valeurs pour ω , et la quantité $m(A)$ peut être interprétée comme la mesure de croyance exactement allouée à $\omega \in A$. L’ignorance complète correspond à $m(\Omega) = 1$ (fonction de masse vide), alors qu’une connaissance parfaite sur la valeur de ω est représentée par l’allocation de la totalité de la masse à un unique singleton de Ω (m est alors appelée une *masse certaine*). Une fonction de masse m peut être représentée de manière équivalente par une fonction de crédibilité *bel*, une fonction de plausibilité *pl*, et une fonction de commonalité *q* définies, respectivement, par :

$$\text{bel}(A) \triangleq \sum_{\emptyset \neq B \subseteq A} m(B) \quad \forall A \subseteq \Omega, \quad (2)$$

$$\text{pl}(A) \triangleq \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A \subseteq \Omega, \quad (3)$$

$$\text{q}(A) \triangleq \sum_{B \supseteq A} m(B) \quad \forall A \subseteq \Omega. \quad (4)$$

Deux masses m_1 et m_2 induites par deux corpus d’évidence sur Ω peuvent être combinées par la règle de combinaison conjonctive. La fonction

de masse résultante $m_1 \odot m_2$ est donnée par :

$$(m_1 \odot m_2)(A) \triangleq \sum_{B \cap C = A} m_1(B) m_2(C) \quad \forall A \subseteq \Omega. \quad (5)$$

Cette règle conjonctive transfère le produit des masses $m_1(B)$ et $m_2(C)$ à l'intersection de B et C . Elle peut aussi s'exprimer à l'aide des commonalités de manière très simple :

$$(q_1 \odot q_2)(A) = q_1(A) q_2(A) \quad \forall A \subseteq \Omega, \quad (6)$$

où $q_1(A)$ et $q_2(A)$ représentent, respectivement, les commonalités associées à m_1 et à m_2 .

Enfin, il faut noter que la théorie des fonctions de croyance permet de travailler sur des cadres de discernement dont la granularité est différente. Soient deux cadres Θ et Ω et une application $\rho : 2^\Theta \rightarrow 2^\Omega$ telle que :

1. l'ensemble $\{\rho(\{\theta\}), \theta \in \Theta\} \subseteq 2^\Omega$ est une partition de Ω ;
2. pour chaque $A \subseteq \Theta$, $\rho(A) = \cup_{\theta \in \Theta} \rho(\{\theta\})$.

L'application ρ est alors appelée raffinement de Θ vers Ω ; par extension, Ω est appelé un raffinement de Θ , et Θ un grossissement de Ω [13]. Lorsque l'on souhaite combiner des informations exprimés sur des grossissement différents de Ω , il faut, avant la combinaison, les ramener sur le même cadre de discernement. Pour transférer une masse $m^\Theta(\cdot)$ d'un grossissement Θ vers Ω , on utilise l'opération d'extension vide [13] définie comme pour tout $B \subseteq \Omega$ par :

$$m^{\Theta \uparrow \Omega}(B) = \begin{cases} m^\Theta(A) & \text{si } B = \rho(A), A \subseteq \Theta, \\ 0 & \text{sinon.} \end{cases} \quad (7)$$

3 Méthode de décomposition par paires de Hüllermeier et al.

Le problème est d'apprendre une application qui associe à chaque instance \mathbf{x} un ordre total $\succ_{\mathbf{x}}$ sur l'ensemble de labels $\mathcal{L} = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$. Un ordre total $\succ_{\mathbf{x}}$ sur \mathcal{L} peut être représenté de manière équivalente par une permutation $\tau_{\mathbf{x}}$ de l'ensemble des entiers $\{1, 2, \dots, m\}$ de telle sorte que $\lambda_i \succ_{\mathbf{x}} \lambda_j \Leftrightarrow$

$\tau_{\mathbf{x}}(i) < \tau_{\mathbf{x}}(j)$ (la valeur $\tau_{\mathbf{x}}(i)$ représente le rang de λ_i dans les préférences de \mathbf{x}). Pour apprendre la fonction d'ordonnancement, on dispose d'un ensemble d'apprentissage, c'est à dire d'un ensemble de couples :

$$\mathcal{D} = \{(\mathbf{x}_1, \tau_1), \dots, (\mathbf{x}_n, \tau_n)\}.$$

L'approche proposée par Hüllermeier et al. [12] consiste à apprendre un classifieur binaire pour chaque couple d'alternatives (i, j) tels que $i < j$. Chaque classifieur \mathcal{M}_{ij} est entraîné à répondre 1 si $\lambda_i \succ \lambda_j$ et 0 dans le cas contraire. Les sorties de chaque classifieur ne sont pas obligatoirement binaires mais plus généralement comprises dans l'intervalle $[0; 1]$. La construction de ces $m(m-1)/2$ classifieurs permet à Hüllermeier de construire une relation floue de préférence $R_{\mathbf{x}}$ associée à chaque entrée \mathbf{x} :

$$R_{\mathbf{x}}(i, j) = \begin{cases} \mathcal{M}_{ij}(\mathbf{x}) & \text{si } i < j, \\ 1 - \mathcal{M}_{ij}(\mathbf{x}) & \text{sinon.} \end{cases}$$

L'ordre total est obtenu simplement en triant les scores $S_{\mathbf{x}}(i)$ calculés pour chaque alternative λ_i de la manière suivante :

$$S_{\mathbf{x}}(i) = \sum_{j \neq i} R_{\mathbf{x}}(i, j). \quad (8)$$

Même si la manière d'ordonner les alternatives semble relativement simpliste, les auteurs montrent de manière intéressante que si les classifieurs binaires fournissent des estimations exactes de la probabilité que $\lambda_i \succ \lambda_j$, alors la procédure proposée minimise un risque calculé à partir du coefficient de corrélation de rang de Spearman.

4 Ordonnancement d'alternatives dans le cadre des fonctions de croyance

4.1 Modèle

Soit \mathcal{P}_m l'ensemble de toutes les permutations possibles de $\{1, 2, \dots, m\}$. Pour chaque couple (i, j) , on choisit d'utiliser un classifieur binaire

évidentiel permettant de décider pour une entrée \mathbf{x} donnée si l'alternative λ_i est préférée ou non à l'alternative λ_j . En réalité, chaque classifieur binaire travaille sur un grossissement particulier Θ_{ij} de \mathcal{P}_m :

$$\Theta_{ij} = \{\theta_{ij}, \overline{\theta_{ij}}\},$$

où θ_{ij} et $\overline{\theta_{ij}}$ désigne des sous-ensembles de permutations définis de la manière suivante :

$$\theta_{ij} = \{\tau \in \mathcal{P}_m / \tau(i) < \tau(j)\},$$

et

$$\overline{\theta_{ij}} = \{\tau \in \mathcal{P}_m / \tau(i) > \tau(j)\}.$$

On suppose que chaque classifieur évidentiel fournit, pour chaque entrée \mathbf{x} , la fonction de masse suivante :

$$\begin{cases} m^{\Theta_{ij}}(\theta_{ij}/\mathbf{x}) = \alpha_{ij} \\ m^{\Theta_{ij}}(\overline{\theta_{ij}}/\mathbf{x}) = \beta_{ij} \\ m^{\Theta_{ij}}(\Theta_{ij}/\mathbf{x}) = 1 - \alpha_{ij} - \beta_{ij}. \end{cases} \quad (9)$$

L'intérêt du classifieur évidentiel est qu'il peut fournir des éléments de preuve en faveur de l'une ou l'autre des alternatives mais qu'il permet aussi d'exprimer le doute.

A partir des masses, on en déduit facilement les plausibilités suivantes :

$$\begin{cases} \text{pl}^{\Theta_{ij}}(\theta_{ij}/\mathbf{x}) = 1 - \beta_{ij}, \\ \text{pl}^{\Theta_{ij}}(\overline{\theta_{ij}}/\mathbf{x}) = 1 - \alpha_{ij}. \end{cases} \quad (10)$$

Notons que, de par la structure particulière de la fonction de masse, ces valeurs correspondent aussi aux commonalités. Ainsi on a :

$$\begin{cases} \text{q}^{\Theta_{ij}}(\theta_{ij}/\mathbf{x}) = 1 - \beta_{ij}, \\ \text{q}^{\Theta_{ij}}(\overline{\theta_{ij}}/\mathbf{x}) = 1 - \alpha_{ij}. \end{cases} \quad (11)$$

4.2 Choix d'un ordre

Au final, compte tenu des informations dont on dispose (c'est à dire des $m^{\Theta_{ij}}(./\mathbf{x})$), il s'agit de déterminer pour une entrée donnée, un ordre total sur les alternatives. Deux méthodes ont été envisagées. La première s'inspire de celle de

Hüllermeier. Nous construisons une relation de préférence floue définie de la manière suivante :

$$R_{\mathbf{x}}(i, j) = \begin{cases} \text{pl}^{\Theta_{ij}}(\theta_{ij}/\mathbf{x}) = 1 - \beta_{ij} & \text{si } i < j, \\ \text{pl}^{\Theta_{ij}}(\overline{\theta_{ij}}/\mathbf{x}) = 1 - \alpha_{ij} & \text{sinon.} \end{cases} \quad (12)$$

L'ordre total final est obtenu en ordonnant les scores $S_{\mathbf{x}}(i)$ calculés de la même manière que Hüllermeier à l'aide de l'équation (8). Cette première approche sera désignée dans ce qui suit comme la "méthode de vote".

La seconde approche considère le problème sous l'angle de la fusion d'information. On dispose de plusieurs sources (les $m(m-1)/2$ classifieurs, exprimés sur des référentiels différents, et l'on cherche une fonction de masse $m^{\mathcal{P}_m}(./\mathbf{x})$ exprimé sur \mathcal{P}_m . Cette fonction de masse s'obtient en étendant chaque fonction de masse sur \mathcal{P}_m par l'opération d'extension vide, puis en combinant de manière conjonctive ces différentes masses :

$$m(./\mathbf{x}) = \bigodot_{i < j} m^{\Theta_{ij} \uparrow \mathcal{P}_m}(./\mathbf{x}). \quad (13)$$

Il est possible de calculer la plausibilité d'une permutation. En effet, si l'on exprime cette combinaison par l'intermédiaire des commonalités (grâce à la relation (6)), on a :

$$\text{q}(./\mathbf{x}) = \prod_{i < j} \text{q}^{\Theta_{ij} \uparrow \mathcal{P}_m}(./\mathbf{x}). \quad (14)$$

Or :

$$\text{q}^{\Theta_{ij} \uparrow \mathcal{P}_m}(\tau/\mathbf{x}) = \begin{cases} 1 - \beta_{ij} & \text{si } \tau(i) < \tau(j) \\ 1 - \alpha_{ij} & \text{sinon.} \end{cases} \quad (15)$$

La commonalité associée à une permutation τ peut donc s'écrire :

$$\text{q}(\tau/\mathbf{x}) = \prod_{\tau(i) < \tau(j)} (1 - \beta_{ij}) \prod_{\tau(i) > \tau(j)} (1 - \alpha_{ij}). \quad (16)$$

où la notation $\tau(i) < \tau(j)$ en indice (respectivement $\tau(i) > \tau(j)$) désigne l'ensemble des couples (i, j) , $i < j$ pour lesquels λ_i est placé avant (respectivement après) λ_j .

La permutation τ étant un élément singleton de \mathcal{P}_m , cette valeur n'est rien d'autre que la plausibilité $pl(\tau/\mathbf{x})$.

Une règle de choix d'un ordre particulier dans \mathcal{P}_m peut consister à chercher la permutation τ^* de plus grande plausibilité. Nous allons voir dans ce qui suit qu'il n'est pas nécessaire d'énumérer toutes les permutations possibles. Pour maximiser $pl(\tau/\mathbf{x})$, il est équivalent de maximiser son logarithme. La permutation optimale s'exprime donc sous la forme suivante :

$$\tau^* = \underset{\tau}{\operatorname{argmax}} \sum_{\tau(i) < \tau(j)} \ln(1 - \beta_{ij}) + \sum_{\tau(i) > \tau(j)} \ln(1 - \alpha_{ij}).$$

La détermination de τ^* peut se faire grâce à la résolution d'un programme linéaire en nombres entiers binaires. Sa formulation est la suivante : on introduit $m(m-1)/2$ variables x_{ij} prenant soit la valeur 0, soit la valeur 1. La valeur 1 correspond au fait de placer le label λ_i avant le label λ_j , la valeur 0 correspond au placement contraire. Pour déterminer les x_{ij} optimaux, on résout le programme linéaire suivant :

$$\max_{x_{ij} \in \{0,1\}} \sum_{i < j} x_{ij} \ln(1 - \beta_{ij}) + \sum_{i < j} (1 - x_{ij}) \ln(1 - \alpha_{ij}). \quad (17)$$

sous les contraintes :

$$\begin{cases} x_{ij} + x_{jk} - 1 \leq x_{ik} & \forall i < j < k, \\ x_{ik} \leq x_{ij} + x_{jk} & \forall i < j < k. \end{cases} \quad (18)$$

Les contraintes permettent d'assurer la cohérence des valeurs x_{ij} entre elles, c'est à dire la transitivité de la relation d'ordre recherchée. La première contrainte assure que si $\tau(\lambda_i) < \tau(\lambda_j)$ et $\tau(\lambda_j) < \tau(\lambda_k)$ alors $\tau(\lambda_i) < \tau(\lambda_k)$ ($x_{ij} = 1, x_{jk} = 1$ alors $x_{ik} = 1$). La seconde contrainte assure que si $\tau(\lambda_i) > \tau(\lambda_j)$ et $\tau(\lambda_j) > \tau(\lambda_k)$ alors $\tau(\lambda_i) > \tau(\lambda_k)$ ($x_{ij} = 0, x_{jk} = 0$ alors $x_{ik} = 0$).

La permutation optimale est entièrement déterminée par les valeurs prises par les x_{ij} :

$$\begin{cases} \tau^*(i) < \tau^*(j) & \text{si } x_{ij} = 1, \\ \tau^*(i) > \tau^*(j) & \text{si } x_{ij} = 0. \end{cases} \quad (19)$$

Tableau 2 – Jeux de données

Nom	# exemples	# attr.	# labels
authorship	1513	70	4
glass	382	9	6
iris	150	4	3
pendigits	19784	16	10
segment	4158	18	7
vehicule	1518	18	4
vowel	944	10	11
wine	314	13	3

On désignera cette méthode dans la partie expérimentale sous le nom de “méthode du maximum de plausibilité”.

5 Résultats expérimentaux

5.1 Jeux de données

Pour tester notre méthode, nous avons utilisé des jeux de données construits par Hüllermeier et disponibles sur sa page web (<http://www.uni-marburg.de/fb12/kebi/research/repository/> labelrankingdata). Les jeux de données proviennent de l'UCI Machine Learning Repository (www.ics.uci.edu/mllearn/). Originellement, ils sont destinés à tester des algorithmes de discrimination dans le cas classique où l'on cherche à prédire un unique label. Hüllermeier les a transformé en problèmes d'ordonnement d'alternatives par la méthode suivante : pour chaque jeu de données, un classifieur de type Naive Bayes est entraîné sur l'ensemble des données. Ensuite, pour chaque exemple, tous les labels sont ordonnés en fonction de leur probabilité *a posteriori*. Une description de ces données est fournie dans le tableau 2.

5.2 Méthodologie

Nous utilisons comme classifieur binaire les k-plus proches voisins (k-PPV) évidentiels de Denoeux [6] (programmes matlab téléchargeables à l'adresse www.hds.utc.fr/~tdenoeux). Pour comparer les deux manières de prédire l'ordonnement des labels (méthode des votes, méthode

Tableau 3 – Résultats expérimentaux ; taux de Kendall obtenus sur les données de test moyennés sur 10 répétitions (écart-type sur 10 répétitions).

Fichier	Vote	Max de Pl.
authorship	0.8679 (0.0105)	0.8687 (0.0110)
glass	0.8066 (0.0347)	0.8080 (0.0349)
iris	0.9156 (0.0354)	0.9218 (0.0299)
pendigits	0.6444 (0.0109)	0.6485 (0.0103)
segment	0.6399 (0.0089)	0.6458 (0.0063)
vehicule	0.8278 (0.0092)	0.8305 (0.0102)
vowel	0.7284 (0.0208)	0.7284 (0.0185)
wine	0.9161 (0.0264)	0.9154 (0.0260)

du maximum de plausibilité), nous procédons de la façon suivante. Chaque ensemble de données est divisé en deux ensembles de taille équivalente, l’un pour l’apprentissage, l’autre pour le test. Pour chaque ensemble d’apprentissage, nous déterminons la valeur optimale de k , le nombre de voisins, par validation croisée (découpage en 10 sous-ensembles). La qualité de la prédiction est alors mesurée sur l’ensemble test grâce à la valeur du tau de Kendall mesuré entre les rangs prédits et les rangs connus. La tau de Kendall est un coefficient de corrélation de rangs qui permet de comparer deux ordre totaux en se fondant sur le nombre d’inversions de paires d’alternatives entre les deux ordres. Cette procédure est répétée 10 fois.

5.3 Résultats

Nous reproduisons dans le tableau 3 les résultats moyens ainsi que l’écart-type sur les 10 répétitions. On constate que la méthode du maximum de plausibilité donne toujours des résultats aussi bons ou meilleurs que la méthode des votes. Notons cependant que seule la différence observée sur le jeu de données `glass` est statistiquement significative (test de student de comparaison de moyennes à 5%). Les résultats sont tout à fait comparables à ceux rapportés dans [12].

5.4 Rejet

La méthode du maximum de plausibilité, grâce au calcul explicite de la plausibilité de la permutation optimale, a l’avantage d’offrir un moyen naturel d’évaluer la confiance que l’on peut placer dans le résultat. Ainsi, il est facile de mettre en place une procédure de rejet qui consiste à ne pas se prononcer si la valeur de plausibilité obtenue est trop faible. A titre d’exemple, on trouve en Figure 1 l’évolution du tau de Kendall moyen en fonction du pourcentage de rejet pour le jeu de données `vehicule` (nombre de plus proches voisins égal à 3). La courbe est obtenue en ordonnant les instances selon leurs valeurs de plausibilité et en écartant un certain pourcentage d’instances de plus faible plausibilité (taux de rejet variant de 0 à 90%). Le tau de Kendall moyen est ensuite calculé sur les instances restantes. La courbe présente les valeurs moyennes de tau de Kendall ainsi que leur écart-type calculés sur 10 répétitions de cette procédure, en choisissant de manière aléatoire les ensembles d’apprentissage et de test. On voit que le rejet, même pour un pourcentage relativement faible de points, permet d’améliorer de manière significative la qualité de la prédiction. On trouve en Figure 2 une courbe similaire qui présente les taux de Kendall moyen et leur écart-type en fonction du seuil de plausibilité choisi.

Notons que la procédure de rejet ne peut être mise en place que si le nombre de voisins est strictement supérieur à 1. En effet, dans le cas du 1-PPV, pour chaque classifieur évidentiel, seule l’un des deux valeurs α_{ij} ou β_{ij} est non nulle, il en résulte que toutes les plausibilités associées à la permutation optimale sont égales à un.

6 Conclusion

Nous avons proposé dans cet article deux approches fondées sur la théorie des fonctions de croyance pour l’ordonnement d’alternatives. Les deux approches procèdent par décomposition en préférences par paires

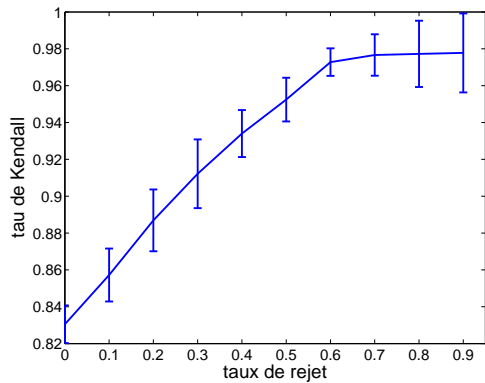


Figure 1 – Courbe qualité de la prédiction-rejet (jeu de données vehicule).

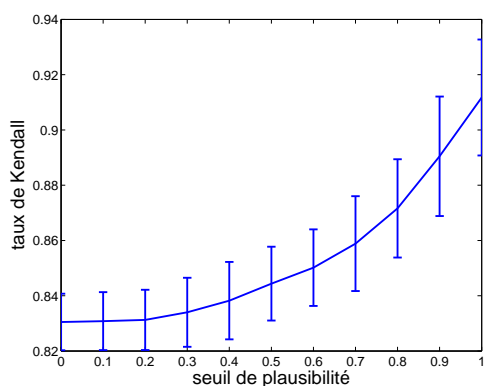


Figure 2 – Courbe qualité de la prédiction-seuil de plausibilité (jeu de données vehicule).

comme il proposé dans [12]. Les méthodes divergent par la manière dont sont combinées les sorties des classifieurs par paires et dont l'ordonnancement final est calculé. La première méthode s'inspire du travail de Hüllermeier et al et repose sur un système de vote des classifieurs. La seconde méthode, qui place le problème sous l'angle de la fusion d'informations, utilise les outils de la théorie des fonctions de croyance pour combiner et réconcilier des informations exprimées sur des référentiels différents. Les premiers résultats expérimentaux ne montrent pas de réelles différences significatives entre les deux méthodes. Cependant, la méthode du maximum de plausibilité offre l'avantage de permettre la mise en place d'une procédure de rejet, grâce à l'exploitation des valeurs de plausibilités. De plus, outre qu'elle nous paraît mieux adaptée

au cadre évidentiel, cette approche nous semble ouvrir la voie à d'autres types de décisions possibles comme l'ordonnancement partiel d'alternatives. Ce point est actuellement à l'étude.

Références

- [1] K. Brinker, and E. Hüllermeier. Case-based label ranking. In *Proceedings of the Seventeenth European Conference on Machine Learning*, 2006.
- [2] W. W. Cohen, and R.E. Schapire, and Y. Singer. Learning to order things. *Journal of Artificial Intelligence Research*, 10,243-270, 1999.
- [3] K. Crammer, and Y. Singer. Pranking with ranking. In *Advances in Neural Information Processing Systems 14*, 2001.
- [4] W. Chu, and Z Ghahramani. Gaussian processes for ordinal regression. *Journal of Machine Learning Research*, 6, 1019-1041, 2005.
- [5] O. Dekel, and C. D. Manning, and Y. Singer. Log-linear models for label ranking. In *Advances in Neural Information Processing Systems 16*, 2003.
- [6] T. Denoeux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(05), 804-813, 1995.
- [7] A. Elisseeff, and J. Weston. A kernel method for multi-labelled classification. In *Advances in Neural Information Processing Systems 14*, 2001.
- [8] Y. Freund, and R. Iyer, and R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, 4, 933-969, 2003.
- [9] J. Fürnkranz, and E. Hüllermeier. Pairwise preference learning and ranking. In *Proceedings of the Fourteenth European Conference on Machine Learning*, 2003.
- [10] S. Har-Peled, and D. Roth, and Dav Zimak. Constraint classification for multiclass classification and ranking. In *Advances in Neural Information Processing Systems 15*, 2002.
- [11] R. Herbrich, and T. Graepel, and K. Obermayer. Large margin rank boundaries for ordinal regression. In Peter J. Bartlett, Bernhard Schölkopf, Dale Schuurmans, and Alex J. Smola, editors, *Advances in Large Margin Classifiers*, 115-132. MIT Press, Cambridge, MA, 2000.
- [12] E. Hüllermeier, and J. Fürnkranz, and W. Cheng, and K. Brinkler. Label ranking by pairwise preference. *Artificial Intelligence*, 172, 1897-1916, 2008.
- [13] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.
- [14] P. Smets, and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66, 191-243, 1994.
- [15] Z. Younes, and F. Abdallah, and T. Denoeux. An Evidence-Theoretic k-Nearest Neighbor Rule for Multi-Label Classification. In *Proceedings of the 3rd International Conference on Scalable Uncertainty Management (SUM 2009)*, September 2009, Washington, DC, USA, LNAI-5785, pp. 297-308, Springer-Verlag.