

# La méthode $RA^kEL$ évidentielle pour la classification multi-label

## The evidential $RA^kEL$ method for multi-label classification

Sawsan Kanj

Fahed Abdallah

Thierry Deneux

Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, France

### Résumé :

La classification multi-label permet d'associer à une observation donnée une ou plusieurs classes simultanément.  $RA^kEL$  est l'une des méthodes utilisées pour répondre à ce problème. Cette méthode est basée sur la division de l'ensemble des classes en plusieurs sous-ensembles, chaque combinaison possible de labels étant considérée comme une nouvelle classe. Chacun de ces sous-ensembles est utilisé pour l'apprentissage d'un classifieur multi-label décidant sur les classes lui appartenant. Pour la décision finale sur l'appartenance d'un individu à une classe, une stratégie de vote est utilisée. Dans cet article, nous présentons une méthode de classification multi-label consistant à utiliser la méthode  $RA^kEL$  conjointement avec la théorie des fonctions de croyance adaptée au formalisme des variables multi-valuées. L'utilisation de la théorie des fonctions de croyance rend possible l'association d'une fonction de masse à chaque classifieur. Ces fonctions sont ensuite combinées par un opérateur adapté dans le but de donner une décision finale sur l'appartenance d'un individu à un ensemble de classes. Plusieurs résultats sur des exemples réels montrent les bonnes performances de cette méthode par rapport à la méthode  $RA^kEL$  de base.

### Mots-clés :

Classification multi-label, fonctions de croyance.

### Abstract:

Multi-label classification deals with problems in which each instance can be associated with a set of labels. An effective multi-label method, named  $RA^kEL$ , randomly breaks the initial set of labels into smaller sets and trains a single-label classifier in each of this subset. To classify an unseen instance, the predictions of all classifiers are combined using a voting process. In this paper, we adapt the  $RA^kEL$  approach under the belief function framework applied to set-valued variables. Using evidence theory allows us to handle partial information by associating a mass function to each classifier and combining them conjunctively. Experiments on real datasets demonstrate that our approach improves classification performances as compared to the basic  $RA^kEL$  method.

### Keywords:

Multi-label classification, belief functions.

## 1 Introduction

Dans les problèmes de classification multi-label, chaque objet peut appartenir simul-

tanément à plusieurs classes, contrairement aux problèmes standards de classification mono-label dans lesquels un objet appartient à une seule classe. Par exemple, dans la catégorisation de textes [4], chaque document peut appartenir à plusieurs thèmes, tels que la santé, le sport et la nutrition. Dans la classification de chansons [5], chaque morceau peut évoquer plusieurs émotions simultanément, etc.

Plusieurs méthodes ont été proposées dans la littérature pour traiter la problématique de l'apprentissage multi-label [6, 8, 9]. Ces méthodes peuvent être divisées en deux groupes selon la façon dont on traite l'ensemble des données d'apprentissage. Le premier groupe de méthodes transforme le problème d'apprentissage multi-label en un ou plusieurs problèmes d'apprentissage mono-label, tandis que le second groupe se base sur l'adaptation directe des algorithmes de classification mono-label pour l'apprentissage multi-label [6].

Une méthode fréquemment utilisée pour la classification multi-label et appartenant au premier groupe de méthodes est la méthode BR (*Binary Relevance*) [7]; cette méthode transforme le problème d'apprentissage initial en  $Q$  problèmes de classification binaire,  $Q$  désignant le nombre de classes possibles. Chaque classifieur binaire est utilisé pour séparer une classe des autres. La méthode BR a l'avantage d'être simple et peu coûteuse en terme de temps de calcul, mais son problème majeur réside dans le fait qu'elle ne tient pas compte des corrélations éventuelles entre les classes.

La méthode dite *label powerset* (LP) appar-

tient à la même catégorie [7]. Pour cette méthode, chaque ensemble de labels existant dans l'ensemble d'apprentissage de taille  $N$  est considéré comme une nouvelle classe. Le problème d'apprentissage multi-label est ensuite transformé en un problème de classification mono-label où le nombre de classes est au plus égal à  $\min(2^Q, N)$ . Pour un nouvel exemple à classifier, la classe (multi-label) la *plus probable* est prédite. La méthode LP tient compte des corrélations entre les classes, mais elle souffre en général des problèmes de complexité algorithmique. Dans [7], les auteurs proposent une variante de la méthode LP. Le principe de cette méthode, appelée  $RAkEL$  (Random  $k$ -labELsets), est de générer des ensembles de classes aléatoires de petite taille par rapport au nombre total des classes ( $Q$ ) et de faire l'apprentissage du classifieur multi-label pour chaque ensemble aléatoire en utilisant la méthode LP. Les classes sont ensuite déterminées par une stratégie de vote en utilisant un seuil à déterminer. La méthode  $RAkEL$  a une complexité inférieure à celle de la méthode LP. Dans cet article nous nous intéressons à la méthode  $RAkEL$ . Dans cette méthode, trois paramètres sont à identifier : le nombre d'ensembles aléatoires (ou, d'une manière équivalente, le nombre de classifieurs), la taille de ces ensembles et le seuil utilisé lors de l'application de la stratégie de vote.

L'objectif de notre travail est de réduire la perte d'information causée par l'utilisation de la méthode  $RAkEL$  (du fait que l'apprentissage de classifieurs est effectué sur un sous-ensemble de labels) tout en tenant compte des corrélations possibles entre les différentes classes. Nous proposons pour cela de maintenir la structure de l'approche  $RAkEL$  et de combiner la sortie de différents classifieurs dans le cadre de la théorie des fonctions de croyance, en utilisant le formalisme de représentation de l'incertitude pour les variables multi-valuées proposé dans [3]. L'analyse discriminante linéaire (ADL) est utilisée comme classifieur multi-classe de base. Les sorties de l'ADL peuvent

être exploitées pour la création des fonctions de masse sur des variables multi-valuées. La méthode développée, nommée  $RAkEL-ADL-Evidentielle$ , présente l'intérêt de ne pas nécessiter de fixer un seuil pour appliquer la stratégie de vote, la décision étant prise après combinaison des fonctions de masse par la règle de Dempster [2, 3].

Cet article est organisé comme suit. Le paragraphe 2 introduit le formalisme des variables multi-valuées et leur représentation en utilisant la théorie des fonctions de croyance de Dempster-Shafer. Nous présentons la méthode  $RAkEL-ADL-Evidentielle$  dans le paragraphe 3. Dans le paragraphe 4, nous comparons la méthode proposée à la méthode  $RAkEL-ADL$ . Enfin, le paragraphe 5 conclut l'article.

## 2 Fonctions de croyance sur des variables multi-valuées

Soit  $Y$  une variable multi-valuée prenant une ou plusieurs valeurs dans un ensemble  $\Omega$ . L'approche directe pour exprimer des connaissances sur  $Y$  consiste à la considérer comme une variable univaluée prenant une valeur dans  $\Theta = 2^\Omega$ . Ainsi, pour représenter des informations imprécises sur  $Y$ , nous avons à manipuler des sous-ensembles de  $\Theta$ . Comme expliqué dans [3], cette approche implique la manipulation des éléments d'un ensemble de cardinalité très élevée, ce qui rend le problème rapidement insoluble dès que la taille de  $\Omega$ , et donc de  $\Theta$ , augmente. Dans [3], un formalisme pour quantifier l'incertitude sur une variable multi-valuée à l'aide des fonctions de croyance a été présenté. Ce formalisme repose sur la définition d'une famille  $\mathcal{C}(\Omega)$  de sous-ensembles de  $2^\Omega$  fermée par intersection et ayant une structure de treillis. Chaque élément de  $\mathcal{C}(\Omega)$  est décrit via une paire de sous-ensembles disjoints de  $\Omega$ . La plupart des notions de la théorie de Dempster-Shafer peut être définie sur  $\mathcal{C}(\Omega)$  au prix d'une augmentation modérée de la complexité par rapport au cas classique des variables mono-valuées. Nous rappelons ci-après

les principes de base du formalisme des fonctions de croyance pour les variables multi-valuées.

Étant donnés deux sous-ensembles disjoints de  $\Omega$ ,  $A$  et  $B$ , soit  $\varphi(A, B)$  le sous-ensemble de  $2^\Omega$  défini par :

$$\varphi(A, B) = \{C \subseteq \Omega / A \subseteq C \subseteq \overline{B}\}. \quad (1)$$

$\varphi(A, B)$  est donc le sous-ensemble de  $\Theta$  composé de tous les sous-ensembles de  $\Omega$  contenant  $A$  et disjoints de  $B$ .

Pour un ensemble fini  $\Omega$ ,  $\mathcal{C}(\Omega)$  est l'ensemble de toutes les parties de  $\Theta = 2^\Omega$  de la forme  $\varphi(A, B)$ , complété par l'ensemble vide de  $\Theta$ , noté  $\emptyset_\Theta$  :

$$\mathcal{C}(\Omega) = \{\varphi(A, B) / A \subseteq \Omega, B \subseteq \Omega, A \cap B = \emptyset_\Omega\} \cup \{\emptyset_\Theta\}. \quad (2)$$

Le cardinal du sous-ensemble  $\mathcal{C}(\Omega)$  de  $\Theta$  est égal à  $3^{|\Omega|} + 1$ . Il est donc largement inférieur au cardinal de  $2^\Theta$  ( $|2^\Theta| = 2^{2^{|\Omega|}}$ ), tout en étant assez riche pour exprimer des croyances dans de nombreuses situations réalistes. L'idée intuitive qui sous-tend cette approche est que, en exprimant la connaissance sur une variable multi-valuée  $\mathbf{Y}$ , il est souvent adéquat de spécifier des ensembles de valeurs qui sont *certainement prises* par  $\mathbf{Y}$ , et d'autres qui *ne sont certainement pas prises* par  $\mathbf{Y}$ .

Le sous-ensemble choisi  $\mathcal{C}(\Omega)$  est stable par intersection, c'est-à-dire que pour tout  $\varphi(A, B)$ ,  $\varphi(A', B') \in \mathcal{C}(\Omega)$ ,  $\varphi(A, B) \cap \varphi(A', B') \in \mathcal{C}(\Omega)$ . L'opérateur d'intersection est défini par :

$$\varphi(A, B) \cap \varphi(A', B') = \begin{cases} \varphi(A \cup A', B \cup B') & \text{si } (A \cup A') \cap (B \cup B') = \emptyset_\Omega, \\ \emptyset_\Theta & \text{sinon.} \end{cases} \quad (3)$$

Notons que le sous-ensemble  $\mathcal{C}$  est muni d'une structure de treillis.

Soit  $m : \mathcal{C}(\Omega) \rightarrow [0, 1]$  une fonction de masse dans  $\mathcal{C}(\Omega)$ . La notation  $m(\varphi(A, B))$  sera simplifiée en  $m(A, B)$ . La fonction de masse  $m$  est appelée *fonction de masse à deux places*. Les fonctions de croyance et de communalité peuvent être calculées à partir de  $m$  en utilisant les formules suivantes :

$$bel(A, B) = \sum_{\varphi(C, D) \subseteq \varphi(A, B)} m(C, D) - m(\Omega, \Omega), \quad (4)$$

$$q(A, B) = \sum_{\varphi(C, D) \supseteq \varphi(A, B)} m(C, D). \quad (5)$$

La somme conjonctive normalisée dans  $\mathcal{C}(\Omega)$  utilisée pour l'agrégation de deux fonctions de masse  $m_1$  et  $m_2$  indépendantes ( $m = m_1 \oplus m_2$ ) est définie comme suit :

$$m(A, B) = \frac{\sum_{\varphi(C, D) \cap \varphi(E, F) = \varphi(A, B)} m_1(C, D) m_2(E, F)}{\sum_{\varphi(C, D) \cap \varphi(E, F) \neq \emptyset_\Theta} m_1(C, D) m_2(E, F)}. \quad (6)$$

Même si l'approche évidentielle présentée ci-dessus réduit le nombre des éléments focaux à  $3^{|\Omega|} + 1$ , la complexité de la méthode reste élevée lorsque la taille de  $\Omega$  est grande. Par exemple, si nous avons un problème multi-label à 20 classes, nous avons à traiter jusqu'à  $3.4868e + 009$  éléments focaux avec l'approche proposée. Pour remédier à ce problème, nous proposons dans la section suivante une méthode basée sur l'approche *RAkEL* dans le but d'appliquer le formalisme évidentielle à des sous-ensembles de  $\Omega$  de tailles réduites, et de combiner les résultats dans le cadre de la théorie des fonctions de croyance.

### 3 Approche évidentielle pour la classification multi-label

Soient  $\mathcal{X} = \mathbb{R}^d$  l'espace d'entrée et  $\Omega = \{\omega_1, \omega_2, \dots, \omega_Q\}$  l'ensemble fini des étiquettes ou classes possibles. Le problème de classification multi-label peut être formulé comme suit. On considère un ensemble d'apprentissage  $\mathcal{D} = \{(\mathbf{x}_1, \mathbf{Y}_1), \dots, (\mathbf{x}_N, \mathbf{Y}_N)\}$  de  $N$  instances de  $\mathcal{X} \times 2^\Omega$  où  $\mathbf{x}_i$  est un vecteur de  $\mathcal{X}$  décrivant l'instance  $i$ , et  $\mathbf{Y}_i \subseteq \Omega$  est l'ensemble des étiquettes ou classes de cette instance. Le but de la classification multi-label est de trouver un classifieur  $\mathcal{H} : \mathcal{X} \rightarrow 2^\Omega$  qui associe à chaque nouvelle observation  $\mathbf{x}$  un ensemble d'étiquettes  $\mathbf{Y}$  :

$$\begin{aligned} \mathcal{H} : \mathcal{X} &\rightarrow 2^\Omega, \\ \mathbf{x} &\rightarrow \mathbf{Y}. \end{aligned} \quad (7)$$

Lorsque le nombre de classes est grand, la mise en oeuvre des méthodes classiques de classifi-

cation multi-label s'avère difficile, même parfois impossible. La méthode RAKEL est une méthode communément utilisée pour répondre à ce problème [7]. Cette méthode est basée sur l'utilisation de sous-ensembles aléatoires  $\Omega_j$ ,  $j = 1, \dots, m$ , de  $\Omega$ . Pour chaque  $\Omega_j$ , un ensemble d'apprentissage, noté  $\mathcal{D}_j$ , est déduit de l'ensemble d'apprentissage initial  $\mathcal{D}$  en remplaçant chaque  $\mathbf{Y}_i$  par son intersection avec  $\Omega_j$ . À l'intérieur de  $\mathcal{D}_j$ , chaque combinaison possible d'étiquettes est considérée comme une nouvelle classe. La méthode ADL est alors utilisée pour l'apprentissage de classifieurs  $h_j$  sur l'ensemble de classes retenues. Notons que la méthode ADL permet de générer un ensemble de fonctions linéaires de taille égale au nombre de classes et permet d'estimer les probabilités a posteriori de chaque groupe de l'ensemble  $\Omega_j$ . Les sorties des classifieurs  $h_j$  permettent la construction des fonctions de masse exprimées dans l'ensemble initial  $\Omega$ . Chacune de ces fonctions aura des ensembles focaux de la forme  $\varphi(A_q, B_q)$ , où  $A_q$  est l'ensemble d'étiquettes inclus dans  $\Omega_j$  et appartenant à l'observation, et  $B_q$  est son complémentaire dans  $\Omega_j$  ( $B_q = \Omega_j \setminus A_q$ ). En considérant ces fonctions de masses comme éléments de croyance dans  $\Omega$ , nous pouvons les combiner en utilisant la règle de combinaison conjonctive (6). Pour prédire l'ensemble des classes  $\hat{Y}$  d'un nouvel individu, nous pouvons procéder de deux manières différentes. La première consiste à comparer deux degrés de croyance :  $bel(\{\omega\}, \emptyset)$  and  $bel(\emptyset, \{\omega\})$  pour chaque étiquette dans  $\Omega$  [3] :

$$\hat{Y} = \{\omega \in \Omega / bel(\{\omega\}, \emptyset) \geq bel(\emptyset, \{\omega\})\}. \quad (8)$$

La deuxième consiste à sélectionner l'ensemble des classes de communalité maximale. Dans [2], les auteurs montrent que ce maximum peut être obtenu en résolvant un problème d'optimisation sous contraintes.

## 4 Résultats expérimentaux

### 4.1 Métriques d'évaluation

L'évaluation d'un système d'apprentissage multi-label se différencie de celle d'un système d'apprentissage mono-label. Pour évaluer la performance de notre classifieur, plusieurs critères existent dans la littérature [6]. Etant donné un ensemble de test  $\mathcal{S} = \{(\mathbf{x}_1, \mathbf{Y}_1), \dots, (\mathbf{x}_M, \mathbf{Y}_M)\}$  et un classifieur  $\mathcal{H}$ , soient  $\hat{Y}_i = \mathcal{H}(\mathbf{x}_i)$  l'ensemble des étiquettes prédites pour  $\mathbf{x}_i$  et  $Y_i$  le vrai ensemble d'étiquettes associées à  $\mathbf{x}_i$ . Nous donnons ci-dessous les définitions de quelques métriques d'évaluation utilisées :

*Coût de Hamming* : Cette métrique évalue combien de fois une paire exemple-classe est mal classée.

$$\mathcal{H}Loss = \frac{1}{M} \sum_{i=1}^M \frac{|Y_i \Delta \hat{Y}_i|}{Q}, \quad (9)$$

où  $\Delta$  est la différence symétrique entre deux ensembles.

*Accuracy* : Cette métrique mesure le degré de similarité entre l'ensemble des classes prédites et le vrai ensemble d'étiquettes :

$$Accuracy = \frac{1}{M} \sum_{i=1}^M \frac{|Y_i \cap \hat{Y}_i|}{|Y_i \cup \hat{Y}_i|}. \quad (10)$$

*Mesure F* : La mesure F correspond à une moyenne harmonique de deux autres métriques : la précision et le rappel. La *précision* est la proportion de vrais positifs sur l'ensemble des classes prédites, tandis que le *rappel* est la proportion des vraies classes qui sont prédites comme positives. D'où :

$$\mathcal{F} = \frac{1}{M} \sum_{i=1}^M \frac{2|Y_i \cap \hat{Y}_i|}{|Y_i| + |\hat{Y}_i|}. \quad (11)$$

La performance est meilleure pour les petites valeurs du *coût de Hamming*, tandis que pour des valeurs élevées des autres métriques traduisent une meilleure performance.

## 4.2 Bases de données utilisées

Pour nos expérimentations, nous avons utilisé deux ensembles de données réelles souvent utilisés dans l'évaluation de méthodes de classification multi-label<sup>1</sup>.

Les données *emotion* sont constituées de 593 morceaux de musique annotés par des juges selon les émotions qui les engendrent : Peur, Joie, Colère, Surprise, Dégoût et Tristesse. Chaque chanson est décrite par 8 rythmes et 64 timbres, soit un total de 72 attributs. En moyenne, le nombre de classes associé à chaque chanson est de 1.868 [5].

Les données *scene* contiennent six catégories de scènes naturelles : Lever/coucher du soleil, Plage/désert, Mer/Paysages maritimes, Montagnes, Verdure/Forêt et Ville. Ces données sont composées de 1211 images d'apprentissage et de 1196 images de test. Chaque image est décrite par  $49 \times 2 \times 3$  attributs. En moyenne, le nombre de classes associé à chaque image est de 1.074 [1].

Le jeu *yeast* contient des données concernant les analyses fonctionnelles des gènes chez la levure *Saccharomyces cerevisiae*. La base d'apprentissage contient 2417 gènes décrits par 103 attributs et regroupés en 14 classes. En moyenne, le nombre de classes associé à chaque gène est de 4.25 [3].

## 4.3 Discussion

Nous comparons les résultats expérimentaux de notre méthode avec ceux de l'approche  $RAkEL$  basée sur la méthode ADL sans l'utilisation de la théorie des fonctions de croyance. Pour les deux méthodes, le nombre de classes  $k$  dans chaque sous-ensemble  $\Omega_j$  est fixé à 3. Pour étudier l'influence du nombre de sous-ensembles  $\Omega_j$  utilisé, nous le faisons varier entre 2 et  $2 \times Q$ , où  $Q$  est le nombre total des classes. Pour la méthode  $RAkEL-ADL$ , nous montrons les résultats pour trois valeurs du seuil

utilisé pour le vote final de l'algorithme.

Étant donné que nous utilisons une génération aléatoire des classes dans chaque sous-ensemble, les résultats peuvent être très sensibles à la combinaison des classes retenues. Nous avons donc calculé la moyenne sur 10 répétitions de chaque algorithme. Les figures 1 à 5 montrent les diagrammes en boîte pour les résultats obtenus sur les trois métriques d'évaluation utilisées. Sur chaque figure, nous montrons les résultats sur les différents jeux de données de la méthode  $RAkEL-ADL$  et pour les trois valeurs du seuil utilisé pour le vote, ainsi que les résultats de la méthode  $RAkEL-ADL-Evidentielle$  avec ses deux variantes de décision. Les tableaux 1 à 3 présentent la moyenne, ainsi que l'écart type des résultats pour les différentes métriques d'évaluation et les trois jeux de données. Ces résultats montrent que, dans la majorité des cas, notre méthode est plus performante que la méthode  $RAkEL-ADL$ . Ces résultats peuvent être justifiés par le fait que cette dernière utilise des sorties de type binaire (0 ou 1) comme réponse de chaque classifieur construit sur un ensemble  $\Omega_j$ , tandis que notre méthode conserve l'information brute sous forme d'une fonction de masse dans le but de combiner les différentes sorties d'une manière efficace. Notre méthode prend donc en compte la totalité de l'information inhérente à chaque classifieur.

Notons qu'en utilisant la méthode  $RAkEL-ADL$  avec une faible valeur pour le seuil de vote, on a tendance à donner une prédiction incluant l'ensemble des vraies classes de l'observation à classer. Ceci explique la supériorité de la méthode  $RAkEL-ADL$  pour un seuil égal à 0.1 en prenant comme critère la *mesure F* sur les données *emotion*. Ces données ont en effet un nombre de classes moyen par observation relativement élevé. Symétriquement, en utilisant une grande valeur de seuil, on a tendance à fournir une prédiction incluse dans l'ensemble de vraies classes de cette observation. Ceci peut expliquer la supériorité de la méthode  $RAkEL-ADL$  pour un seuil égal à 0.9 pour le *coût de*

1. <http://mulan.sourceforge.net/datasets.html>

*Hamming* sur les données *scene* qui ont un nombre de classes moyen par observation très faible (tout juste supérieur à 1).

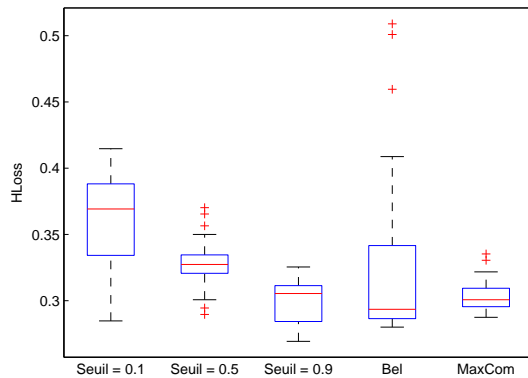


Figure 1 – Diagrammes en boîte du résultat de la métrique coût de Hamming calculée.

## 5 Conclusion

Dans cet article, nous avons présenté une variante de la méthode  $RA^kEL$  pour la classification multi-label. La méthode proposée utilise la théorie des fonctions de croyance pour représenter les sorties des classifieurs, chacun étant construit en considérant un sous-ensemble du cadre de discernement  $\Omega$ . Les fonctions de masse sont définies dans un formalisme développé dans [3] et adapté aux variables multi-valuées. La combinaison conjonctive est utilisée pour obtenir une fonction de masse finale. A partir de celle-ci, nous avons proposé deux méthodes pour la prise de décision finale. La méthode développée a été testée sur deux jeux de données communément utilisés pour l'évaluation des méthodes de classification multi-label. Les résultats obtenus montrent l'intérêt et l'efficacité de la méthode par rapport à la méthode  $RA^kEL$  classique.

## Références

[1] M.R. Boutell, J. Shen, and C.M. Brown. Learning multi-label scene classification.

- Pattern Recognition*, 37 : (9) :1757–1771, 2004.
- [2] T. Denoeux and M.-H. Masson. Evidential reasoning in large partially ordered sets - application to multi-label classification, ensemble clustering and preference aggregation. *Annals OR*, 195(1) :135–161, 2012.
- [3] T. Denoeux, Z. Younes, and F. Abdallah. Representing uncertainty on set-valued variables using belief functions. *Artificial Intelligence*, 174 :479–499, 2010.
- [4] R. Schapire and Y. Singer. Boostexter : a boosting-based system for text categorization. *Machine Learning*, 39 :135–168, 2000.
- [5] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas. Multilabel classification of music into emotions. In *Proc. 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pp. 325-330, 2008.
- [6] G. Tsoumakas and I. Katakis. Multi-label classification : An overview. *International Journal of Data Warehousing and Mining*, 3(3) :1–13, 2007.
- [7] G. Tsoumakas, I. Katakis, and I. Vlahavas. Random k-Labelsets for Multi-Label Classification. *IEEE Transactions on Knowledge and Data Engineering*, 23(7) :1079–1089, 2011.
- [8] G. Tsoumakas, I. Katakis, and I. Vlahavas. Mining multi-label data. In *Data Mining and Knowledge Discovery Handbook*, Springer, 2nd edition, 2010.
- [9] Z. Younes, F. Abdallah, T. Denoeux, and H. Snoussi. A dependent multilabel classification method derived from the k-nearest neighbor rule. *EURASIP Journal on Advances in Signal Processing*, 2011. Article ID 645964, 14 pages, doi :10.1155/2011/645964.

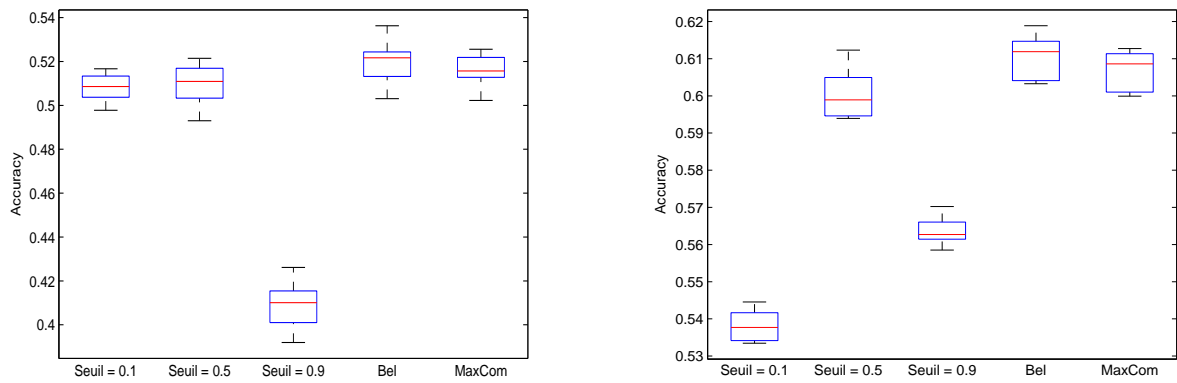


Figure 2 – Diagrammes en boîte des résultats de la métrique *Accuracy* calculée pour les méthodes *RAkEL-ADL* avec des seuils 0.1, 0.5 et 0.9, et *RAkEL-ADL-Evidentielle* avec les deux variantes de décision : bel et maximum de communalité. Figure de droite : données *emotion* ; figure de gauche : données *scene*.

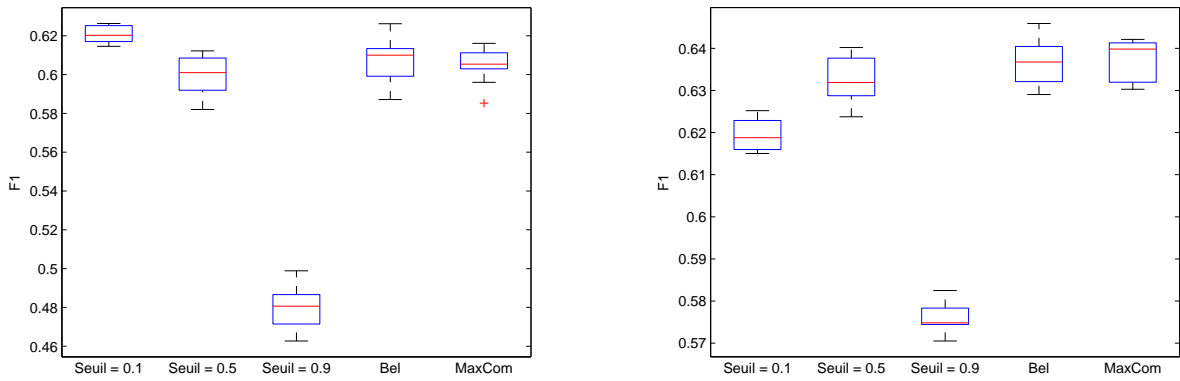


Figure 3 – Diagrammes en boîte des résultats de la mesure *F* calculée pour les méthodes *RAkEL-ADL* avec différents seuils 0.1, 0.5 et 0.9, et *RAkEL-ADL-Evidentielle* avec les deux variantes de décision : bel et maximum de communalité. Figure de droite : données *emotion* ; figure de gauche : données *scene*.

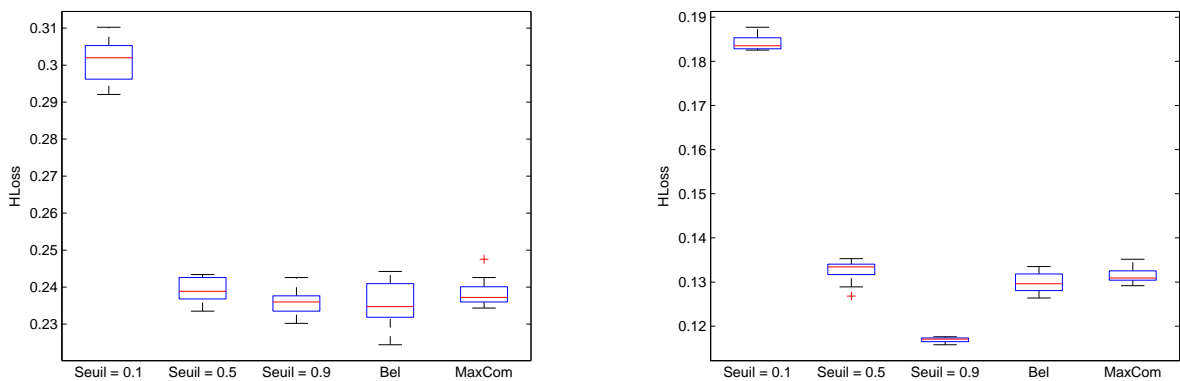


Figure 4 – Diagrammes en boîte des résultats de la métrique *coût de Hamming* calculée pour les méthodes *RAkEL-ADL* avec différents seuils 0.1, 0.5 et 0.9, et *RAkEL-ADL-Evidentielle* avec les deux variantes de décision : bel et maximum de communalité. Figure de droite : données *emotion* ; figure de gauche : données *scene*.

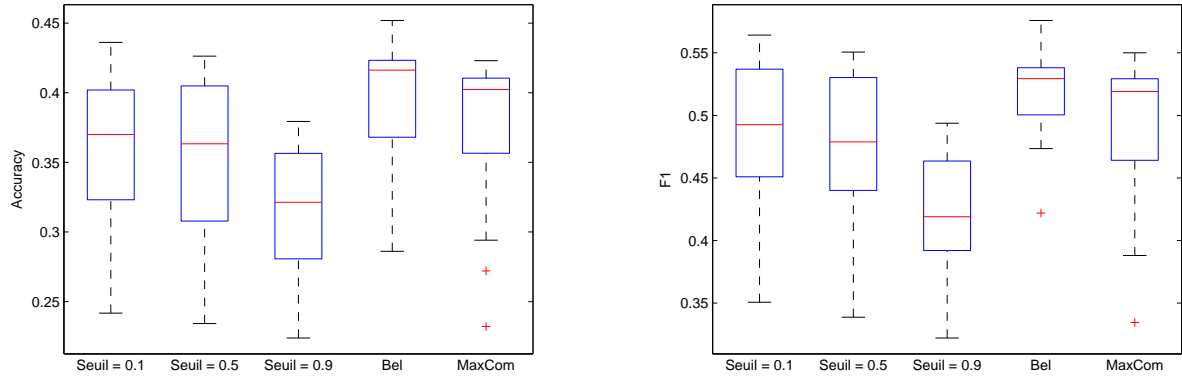


Figure 5 – Diagrammes en boîte des résultats des métriques *Accuracy* et *mesure F* calculées sur les données *yeast* pour les méthodes *RAkEL-ADL* avec différents seuils 0.1, 0.5 et 0.9, et *RAkEL-ADL-Evidentielle* avec les deux variantes de décision : bel et maximum de communalité.

	<i>Rakel-ADL</i> Seuil = 0.1	<i>Rakel-ADL</i> Seuil = 0.5	<i>Rakel-ADL</i> Seuil = 0.9	<i>E-Rakel-ADL</i> <i>Bel</i>	<i>E-Rakel-ADL</i> <i>Max de com</i>
<i>Accuracy</i>	$0.508 \pm 0.006^{(4)}$	$0.509 \pm 0.009^{(3)}$	$0.409 \pm 0.009^{(5)}$	$0.519 \pm 0.009^{(1)}$	$0.516 \pm 0.007^{(2)}$
$F_1$	$0.621 \pm 0.004^{(1)}$	$0.598 \pm 0.011^{(4)}$	$0.479 \pm 0.011^{(5)}$	$0.607 \pm 0.012^{(2)}$	$0.605 \pm 0.009^{(3)}$
<i>HLoss</i>	$0.301 \pm 0.006^{(5)}$	$0.239 \pm 0.003^{(4)}$	$0.236 \pm 0.003^{(2)}$	$0.235 \pm 0.006^{(1)}$	$0.238 \pm 0.004^{(3)}$

Tableau 1 – Résultats numériques (moyenne  $\pm$  écart-type) des différentes métriques d'évaluation pour les données *emotion*.

	<i>Rakel-ADL</i> Seuil = 0.1	<i>Rakel-ADL</i> Seuil = 0.5	<i>Rakel-ADL</i> Seuil = 0.9	<i>E-Rakel-ADL</i> <i>Bel</i>	<i>E-Rakel-ADL</i> <i>Max de com</i>
<i>Accuracy</i>	$0.538 \pm 0.004^{(5)}$	$0.601 \pm 0.006^{(3)}$	$0.564 \pm 0.004^{(4)}$	$0.611 \pm 0.005^{(1)}$	$0.607 \pm 0.005^{(2)}$
$F_1$	$0.612 \pm 0.004^{(4)}$	$0.632 \pm 0.006^{(3)}$	$0.576 \pm 0.004^{(5)}$	$0.636 \pm 0.005^{(2)}$	$0.637 \pm 0.005^{(1)}$
<i>HLoss</i>	$0.184 \pm 0.002^{(5)}$	$0.132 \pm 0.003^{(4)}$	$0.117 \pm 0.001^{(1)}$	$0.129 \pm 0.002^{(2)}$	$0.131 \pm 0.002^{(3)}$

Tableau 2 – Résultats numériques (moyenne  $\pm$  écart-type) des différentes métriques d'évaluation pour les données *scene*.

	<i>Rakel-ADL</i> Seuil = 0.1	<i>Rakel-ADL</i> Seuil = 0.5	<i>Rakel-ADL</i> Seuil = 0.9	<i>E-Rakel-ADL</i> <i>Bel</i>	<i>E-Rakel-ADL</i> <i>Max de com</i>
<i>Accuracy</i>	$0.378 \pm 0.011^{(3)}$	$0.323 \pm 0.011^{(4)}$	$0.319 \pm 0.048^{(5)}$	$0.611 \pm 0.005^{(1)}$	$0.607 \pm 0.005^{(2)}$
$F_1$	$0.512 \pm 0.011^{(4)}$	$0.525 \pm 0.010^{(3)}$	$0.427 \pm 0.050^{(5)}$	$0.636 \pm 0.005^{(2)}$	$0.637 \pm 0.005^{(1)}$
<i>HLoss</i>	$0.395 \pm 0.012^{(4)}$	$0.401 \pm 0.004^{(5)}$	$0.301 \pm 0.014^{(3)}$	$0.129 \pm 0.002^{(1)}$	$0.131 \pm 0.002^{(2)}$

Tableau 3 – Résultats numériques (moyenne  $\pm$  écart-type) des différentes métriques d'évaluation pour les données *yeast*.