

---

# Une nouvelle méthode d'élagage d'ensemble de classifieurs basée sur le concept de marge

**Li Guo, Samia Boukir**

*Institut EGID, Laboratoire G&E, Université de Bordeaux,  
1 allée F. Daguin, F-33670 Pessac, France  
{li.guo,samia.boukir}@egid.u-bordeaux3.fr*

---

*RÉSUMÉ. Les méthodes d'ensemble ont été utilisées avec succès comme schéma de classification. Les algorithmes d'élagage d'ensembles de classifieurs sont apparus afin de réduire la complexité de ce paradigme populaire d'apprentissage. Cet article présente une nouvelle méthode efficace d'élagage d'ensembles qui, non seulement réduit de manière significative la complexité des méthodes d'ensemble, mais permet aussi une meilleure précision de classification que la version sans élagage. Cet algorithme consiste à ordonner tous les classifieurs de base par rapport à leur entropie qui exploite une nouvelle version de la marge des méthodes d'ensemble. La confrontation de cette méthode avec l'approche naïve d'élagage aléatoire des classifieurs de base et avec un autre algorithme d'élagage par ordonnancement a permis de montrer sa supériorité à travers une analyse empirique conséquente.*

*ABSTRACT. Ensemble methods have been successfully used as a classification scheme. The reduction of the complexity of this popular learning paradigm motivated the appearance of ensemble pruning algorithms. This paper presents a new efficient ensemble pruning method which not only highly reduces the complexity of ensemble methods but also performs better than the non-pruned version in terms of classification accuracy. This algorithm consists in ordering all the base classifiers with respect to their entropy which exploits a new version of the margin of ensemble methods. Confrontation with both the naive approach of randomly pruning base classifiers and another ordered-based pruning algorithm turned out convincing in an extensive empirical analysis.*

*MOTS-CLÉS : classification multiple, méthode d'ensemble, élagage d'ensemble, marge, arbre de décision.*

*KEYWORDS: multiple classification, ensemble method, ensemble pruning, margin, decision tree.*

---

DOI:10.3166/TS.27.491-514 © 2010 Lavoisier, Paris

## Extended abstract

Typically, ensemble methods consist of two stages : the production of multiple learners and their combination. We have considered in this work an additional intermediate stage that deals with the reduction of the ensemble size prior to combination : *ensemble pruning*.

We propose an innovative ensemble pruning method based on the margin paradigm. It is an ordering based pruning approach, thus with a lower complexity compared to other state-of-the-art approaches. An alternative definition for the margin, which does not require the true class labels of the instances, is at the core of this algorithm. Like in *boosting*, low margin instances play a key role in building efficient ensemble classifiers. However, our method is not an ensemble building strategy at data level but at classifier level instead, the focus being on the selection of the right classifiers for the best ensemble.

The selection of an optimal sub-ensemble from a given initial bagging ensemble is a difficult combinatorial optimization problem. Only approximate solutions are accessible for ensembles of realistic size. Our objective is to design a pruning technique that reduces the memory requirements and increases the speed of classification of ensembles while maintaining or improving their generalization performance.

We present an efficient and successful *ordering-based* pruning method which uses *entropy* criteria to select the components of the sub-ensemble. The final members of classifiers from the obtained ranking are dynamically selected based on the classification performance of ensembles of increasing size. Thus, it is a more flexible approach than the use of a fixed percentage of learners, and sacrifices some efficiency for more effectiveness.

An innovative margin based entropy, involving a new definition of the margin concept, is introduced. This criterion concentrates on the smaller margin instances which are closer to the class boundaries and thus carry potentially more information about the classes. Furthermore, the class boundaries are the most challenging and thus are usually not well classified.

Our ordering-based ensemble pruning method relies on measuring the entropy of each base classifier in the ensemble. This method is one of *overproduce-and-choose* methods which can be divided into the overproduction and choose phases. In overproduction phases, adequate classifiers are constructed in large numbers, thus overproducing them. In choose phases, firstly, we use these overproduced classifiers to classify a pruning set and calculate the margin of each instance in this set. Then, our margin based entropy is computed for each classifier. All classifiers are then ordered according to their entropy thus leading to a sorted list of classifiers of decreasing entropy with potentially a decreasing reliability. This means that classifiers that correctly predict instances that have a low margin – i.e. the difficult instances – will be highly ranked and included in the pruned ensemble. The first-rank classifier has then the highest entropy. Only the well classified instances, by the base classifiers, are considered in the computation of the entropy. Finally, we choose the first ordered  $M$  classifiers to compose a pruned ensemble.  $M$  is determined through the optimization of an evaluation function, from the pruning set. This function is one

or some of criteria in machine learning such as overall accuracy, margin and diversity.

Unlike other ordering-based methods where each classifier is independently evaluated, our algorithm uses a more global evaluation through the margin-based entropy. Indeed, this criterion involves instance margin values that result from a majority voting of the whole ensemble. Thus, our pruning technique is not only based on individual properties of the ensemble members (e.g. the accuracy of the individual learners). It also takes into account some form of complementarity of the classifiers.

Our new entropy deliberately favors the classifiers with better performance in classifying low margin instances. Thus, it is a boosting-like strategy which aims to increase the performance on low margin instances. Therefore, our strategy of selection of the sub-ensemble of learners will lead to the subset of classifiers with a potentially improved capability to classify complex data in general, and border data in particular. Consequently, it will induce a selection of a subset of learners that are designed to efficiently handle minor classes.

To evaluate our entropy ordering-based ensemble pruning scheme, we use bagging to create an ensemble, and the base classifier is Classification and Regression Trees (CART). Each CART is pruned by just changing the minimum size of terminal nodes. The evaluation function of the best sub-ensemble selection is overall accuracy. We ran experiments on 10 representative data sets from the UCI repository and an airborne urban image. In this empirical analysis, our ensemble pruning method is confronted with both the naive approach of randomly pruning base classifiers and another ordering-based pruning algorithm, namely the accuracy ordering-based ensemble pruning method. It is also compared to the complete ensemble bagging approach, in which no pruning takes place.

Our experimental results show that our method not only largely reduces the complexity of ensemble methods but also performs better than the non-pruned version in terms of classification accuracy especially in case of difficult classes. This approach also outperforms the accuracy ordering based pruning method. Our algorithm is among the fastest methods for ensemble pruning, with a time complexity linear in the size of the ensemble of classifiers. Besides, our algorithm has a potential capability for classifying imbalanced data since it exploits low-margin instances.

## 1. Introduction

La méthode d'ensemble (Dietterich, 1997 ; Dietterich, 2000) est un paradigme populaire d'apprentissage automatique qui construit un modèle de classification en intégrant des composants d'apprentissage multiples. Typiquement, on peut construire un ensemble de classifieurs à quatre niveaux différents (Kuncheva, 2004) : niveau des données (Breiman, 1996 ; Freund *et al.*, 1997), niveau des attributs (Ho, 1998), niveau du classifieur et niveau de combinaison (Kuncheva, 2002). Au niveau des données, les jeux de données sont modifiés puis utilisés pour entraîner chaque classifieur de l'ensemble. Le classifieur de base devrait être instable, autrement dit, de légers changements dans la base d'apprentissage vont entraîner des modifications importantes de la sortie du classifieur. Les réseaux de neurones et les arbres de

décision sont deux exemples typiques de classifieurs instables. L'ensemble de classifieurs peut être aussi construit au niveau des attributs, en utilisant soit des ensembles d'attributs disjoints ou qui se chevauchent. Différents sous-ensembles d'attributs sont générés pour construire chaque classifieur (Ho, 1998). Au niveau classifieur, on se focalise sur le choix du nombre et du type de classifieurs à intégrer dans le système de classification multiple. Finalement, au niveau combinaison, on recherche différentes façons de combiner les classifieurs de base, par exemple par un vote majoritaire.

Le *bagging* (Breiman, 1996) et le *boosting* adaptatif (*adaptive boosting* ou *adaBoost*) (Freund *et al.*, 1997) sont les méthodes d'association de classifieurs, au niveau des données, les plus utilisées et les plus performantes. Le *bagging* est l'acronyme de *bootstrap aggregating*, soit une agrégation de classifieurs de type *bootstrap* (Efron *et al.*, 1994). C'est une approche pour construire des ensembles de classifieurs, chaque classifieur étant entraîné sur une réplique différente de la base d'apprentissage. Les différentes répliques de la base d'apprentissage sont obtenues par *bootstrap*. Chaque échantillon *bootstrap* est obtenu par tirage aléatoire avec remise dans l'échantillon original. L'idée de base réside dans le fait que les différentes répliques de la base d'apprentissage sont légèrement différentes de l'originale mais suffisamment diverses pour obtenir des classifieurs différents qui vont pouvoir être combinés. Le *bagging* utilisant toujours le même poids pour toutes les instances, il ne change pas la distribution des instances, et toutes les classes dans l'algorithme du *bagging* ont le même poids dans la procédure de vote.

L'algorithme *AdaBoost*, le plus utilisé pour implémenter le *boosting*, cherche à produire des classifieurs très précis en combinant les instances problématiques d'un classifieur donné. Les classifieurs résultants sont combinés par un vote pondéré qui permet de tenir compte de la précision de chaque classifieur. Plus précisément, cet algorithme consiste à reproduire de façon itérative la base d'apprentissage pour chaque classifieur en augmentant le poids des instances qui ont été incorrectement classifiées par le classifieur précédent. Ainsi, il se concentre sur les instances les plus difficiles. Dans la plupart des cas, la méthode *AdaBoost* s'avère plus précise que le *bagging*. Cependant, elle a plusieurs inconvénients : sa lourdeur en temps de calcul, le risque de sur-apprendre (*overtraining*) et sa sensibilité au bruit et aux points aberrants (*outliers*), en particulier dans le cas de données en nombre limité (Bauer *et al.*, 1999).

Les méthodes d'ensemble ont été utilisées avec succès dans de nombreux domaines tels que la médecine (Mangiameli *et al.*, 2004), la bio-informatique (Pang *et al.*, 2008), la télédétection (Chehata *et al.*, 2009 ; Gislason *et al.*, 2006), l'écologie (Cutler *et al.*, 2007) et la recherche d'images (Tao *et al.*, 2006). Cependant, elles ont un inconvénient important : la complexité en espace mémoire nécessaire pour stocker les paramètres des classifieurs de l'ensemble, et le temps de calcul requis pour produire une classification, croissent linéairement avec le nombre de classifieurs de l'agrégation (Martínez-Munõz *et al.*, 2009). De plus, il n'est pas toujours vrai que *plus un ensemble est de taille importante mieux c'est* (Zhou *et al.*, 2002a 2002). Toutes ces raisons ont motivé l'apparition d'algorithmes d'élagage (ou *pruning*) d'ensembles (Tsoumakas *et al.*, 2009). Le challenge de la procédure d'élagage est de réduire le nombre de composants de l'ensemble tout en maintenant sa performance. Il s'agit de trouver un sous-ensemble des membres du classifieur multiple qui puisse

réduire la complexité de tout l'ensemble et être aussi performant en termes de classification, voire meilleur que l'ensemble original (sans élagage) (Zhang *et al.*, 2006). C'est une stratégie de construction d'ensembles, au niveau classifieur, qui a déjà fait ses preuves. Cependant, il faudrait aborder le problème de l'élagage avec soin pour gagner réellement en efficacité. Par conséquent, la complexité de ces algorithmes devrait être considérée de près pour choisir à bon escient le plus adéquat pour atteindre ses objectifs de gain en efficacité, voire en qualité de classification. Un bref état de l'art des approches d'élagage d'ensemble les plus utilisées est présenté plus loin.

Nous proposons une méthode d'élagage d'ensembles innovante basée sur le concept de marge des méthodes d'ensemble. Le fait que ce soit la marge d'une classification plutôt que l'erreur brute d'apprentissage qui compte est devenu un facteur clé dans l'analyse et l'utilisation de classifieurs ces dernières années (Bartlett *et al.*, 2000). Ainsi, les exemples durs sur lesquels se focalise le *boosting* sont ceux qui ont la marge la plus faible (Freund *et al.*, 1997). Notre approche est de type ordonnancement (*ordering*), et donc de complexité inférieure à celle des autres approches existantes (Martínez-Munõz *et al.*, 2009). Comme dans le *boosting*, les instances de faible marge jouent un rôle majeur dans l'élaboration de notre système de classifieurs. Cependant, il ne s'agit pas d'une stratégie de manipulation des données d'apprentissage (construction d'ensemble au niveau des données) mais d'une stratégie de sélection de classifieurs (construction d'ensemble au niveau classifieur). Notre technique sera confrontée à une autre approche de même catégorie, à savoir la méthode d'élagage basée sur l'ordonnancement par précision (*accuracy-ordered*). Une analyse comparative avec l'approche naïve d'élagage aléatoire de classifieurs de base va être aussi effectuée.

Cet article est organisé de la façon suivante. La marge des méthodes d'ensemble est abordée dans la section suivante et présente une nouvelle définition de ce concept clé. La section 3 décrit brièvement les méthodes existantes en élagage d'ensembles. Notre méthode d'élagage d'ensembles, basée sur l'ordonnancement par entropie, qui intègre notre propre version de la marge des ensembles, est alors introduite en section 4. La validation de notre approche présentée en section 5 est consacrée à nos résultats expérimentaux. Dans cette partie, la confrontation de notre méthode aussi bien avec l'approche naïve d'élagage aléatoire de classifieurs de base qu'avec un autre algorithme classique d'élagage par ordonnancement sera interprétée. Une conclusion et des perspectives sont enfin données en section 6.

## 2. Marge des méthodes d'ensemble

### 2.1. Théorie

Le concept de marge des méthodes d'ensemble a été initialement appliqué à la compréhension et à l'évaluation des techniques d'agrégation et de coopération de classifieurs. Selon la définition de Schapire (Schapire *et al.*, 1998), en s'inspirant des notations utilisées dans les méthodes à noyaux, la marge d'une instance  $x$  s'exprime par l'équation [1], où  $v_y$  représente le nombre de votes pour la classe réelle  $y$  et  $v_c$  le nombre de votes pour toute autre classe  $c$ . Les valeurs de la marge se situent dans l'intervalle  $[-1, 1]$ . Une valeur de marge positive indique que l'instance

correspondante a été bien classifiée alors qu'une valeur négative révèle qu'elle a été, au contraire, mal classifiée. Plus la marge est importante, plus la classification associée est fiable *a priori*.

$$marge(x, y) = \frac{v_y - \max_{c=1, \dots, L \cap c \neq y} (v_c)}{\sum_{c=1}^L (v_c)} \quad [1]$$

En outre, la marge d'une instance révèle certaines propriétés de celle-ci. Une grande valeur positive signifie que la plupart des classifieurs de base de l'ensemble ont classifié correctement cette donnée, ce qui implique que cette dernière se situe au centre de la distribution de tous les membres de la classe associée ou aux alentours du centre. Ce type d'instances porte des informations générales sur la classe correspondante. En revanche, une valeur négative importante montre que seule une petite partie des classifieurs de base ont correctement classifié l'échantillon correspondant, ce qui indique que ce type de données représente probablement du bruit ou des points aberrants (*outliers*) dans la classe associée. Une valeur proche de 0 reflète le fait que le nombre de classifieurs de base qui ont bien classifié cette instance et le nombre de classifieurs de base qui ont associé une autre étiquette de classe à cet élément sont proches. Cette situation indique que l'instance est probablement en bordure de classes candidates. Ce type de données porte des informations spécifiques sur ces classes.

## 2.2. Une nouvelle définition de la marge

Dans un problème de classification, il paraît judicieux de porter plus d'intérêt sur les instances en bordure de classes parce qu'elles portent plus d'informations significatives sur les classes. Dans ce cas, les vraies étiquettes de classe de ces instances ne sont pas indispensables. Pour mettre en valeur ces données particulières, nous proposons une nouvelle définition de la marge (Guo *et al.*, 2010) qui s'exprime par l'équation [2], où  $c_1$  est la classe de vote majoritaire pour l'instance  $x$  et  $v_{c_1}$  le nombre de votes attribués à cette classe,  $c_2$  est la seconde classe la plus populaire et  $v_{c_2}$  le nombre de votes correspondant. Notre marge prend des valeurs entre 0 et +1. Cette alternative à la définition classique de la marge présente une propriété intéressante : *les valeurs les plus petites indiquent les exemples les plus difficiles à classifier*. Plus la marge est faible, plus l'instance correspondante est *a priori* proche de la frontière des classes, et plus la quantité d'informations qu'elle fournit est importante. En outre, notre concept de marge ne requiert pas au préalable la vraie étiquette de classe de l'instance  $x$ . Par conséquent, elle est potentiellement plus robuste au bruit puisqu'elle n'est pas affectée par des erreurs d'étiquetage de classe qui peuvent apparaître dans la base d'apprentissage elle-même.

Pour montrer le lien entre la valeur de marge de chaque instance et les caractéristiques de cette dernière, nous avons utilisé le *bagging* pour créer un ensemble avec, comme classifieur de base, l'arbre de classification et de régression ou *Classification and Regression Tree* (CART) (Breiman *et al.*, 1984), pour classifier un jeu de données synthétiques *Sin*. La figure 1 montre clairement que la majorité des instances de faible marge de ce jeu de données se trouve sur le bord de chaque classe, se comportant ainsi comme des vecteurs de support (ou *Support Vectors*), à la base

de l'approche SVM, que nous utilisons dans notre méthode d'élagage d'ensemble.

$$\begin{aligned} \text{marge}(x) &= \frac{v_{c_1} - v_{c_2}}{\sum_{c=1}^L (v_c)} \\ &= \frac{\max_{c=1, \dots, L} (v_c) - \max_{c=1, \dots, L \cap c \neq c_1} (v_c)}{T} \end{aligned} \quad [2]$$

où  $T$  représente le nombre d'arbres de décision (CART) de l'ensemble.

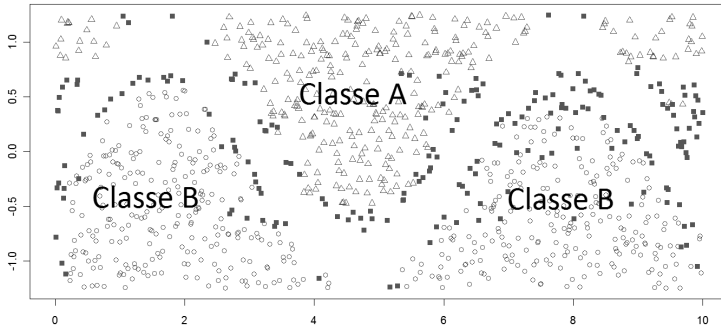


Figure 1. Instances de faible marge (affichées par des carrés pleins) du jeu de données Sin

Nous avons déjà utilisé avec succès cette nouvelle définition de la marge des méthodes d'ensembles pour la sélection de vecteurs de support (*Support Vectors*) visant à réduire la complexité de la phase d'apprentissage des machines à vecteurs de support (*Support Vector Machines* ou SVM) (Guo *et al.*, 2010a). Les instances de plus faible marge ont été sélectionnées comme vecteurs de support. Cette stratégie de sélection a permis de réduire la taille de la base d'apprentissage de façon très significative sans affecter la précision de classification des classifieurs SVM résultants.

D'autre part, nous l'avons utilisée afin d'améliorer la classification de classes mineures et de jeux de données déséquilibrés (*imbalanced data*) (Guo *et al.*, 2010b). Nous avons en effet développé un schéma de classification à deux passes, par forêts aléatoires (*random forests*). Notre stratégie de classification consiste à affiner, après une première classification, la classification des instances de faible marge et seulement celles-ci, les instances de plus forte marge étant considérées comme étant *a priori* bien classifiées. Le second classifieur est entraîné avec une majorité d'instances de faible marge, afin d'avoir une meilleure capacité à traiter les cas les plus difficiles (de faible marge donc). Nous combinons ainsi un classifieur de forte marge avec un classifieur de faible marge.

Dans ce travail, nous allons exploiter la marge des exemples pour évaluer la pertinence de chaque classifieur de base afin d'améliorer la performance globale du sous-ensemble les associant.

### 3. Élagage d'ensemble

Avant de décrire brièvement les principales méthodes d'élagage (*pruning*) d'ensembles existantes, nous introduisons quelques conventions de notation. L'ensemble original de  $T$  classifieurs de base  $C_t$  est noté  $C = (C_t, t = 1, 2, \dots, T)$ . La fonction d'évaluation d'une méthode d'élagage considère un ensemble de  $N$  données, constituant l'échantillon d'élagage (*pruning set*) désigné par  $V = ((x_i, y_i), i = 1, 2, \dots, N)$ , où  $x_i$  est un vecteur de valeurs d'attributs et  $y_i$  est la valeur d'étiquette de classe, qui peut être inconnue.  $|y|$  représente le nombre de classes de la tâche de classification.

Traditionnellement, on crée autant de classifieurs de base que possible pour constituer un ensemble. Il est vrai que pour obtenir une grande précision de classification qui soit stable, un certain nombre de classifieurs de base est requis pour construire l'ensemble. Cependant, plus on crée de classifieurs de base, plus la probabilité d'obtenir des classifieurs de base similaires est forte, moins on produit de diversité (Kuncheva *et al.*, 2003 ; Brown *et al.*, 2005 ; Banfield *et al.*, 2005), et plus on gaspille de ressources. Evidemment, créer un ensemble entier de classifieurs de base similaires n'est pas dans la philosophie de l'approche de construction d'ensembles qui vise plutôt à combiner des classifieurs divers pour engendrer la classification multiple la plus fiable. Ainsi, il n'est pas toujours vrai que *plus un ensemble est de taille importante mieux c'est*. Par conséquent, la conception d'un modèle de classification, intégrant des composants d'apprentissage multiples, devrait être abordée avec précaution pour engendrer une réelle efficacité.

L'élagage d'ensemble est une stratégie de sélection qui se concentre sur la recherche d'un bon sous-ensemble de classifieurs qui soit tout aussi, voire plus efficace, que l'ensemble original (Zhang *et al.*, 2006). Malheureusement, la complexité de la recherche du meilleur sous-ensemble d'un ensemble original de taille  $T$  est en  $O(2^T)$ . C'est un problème NP-complet (Garey *et al.*, 1979), dont la complexité croît de façon exponentielle en fonction de  $T$ . De plus, il faudrait construire initialement un ensemble de grande taille pour avoir le plus de chance d'en déduire le meilleur sous-ensemble possible. Cependant, le recours à une méthode d'élagage très complexe pour simplifier un ensemble de classifieurs perdrait tout son intérêt. Une stratégie d'élagage d'ensembles plausible devrait alors non seulement réduire la taille de l'ensemble, mais aussi conduire à une meilleure précision que l'ensemble original, tout en étant aussi simple que possible.

De nombreuses méthodes d'élagage d'ensembles ont été proposées dans la littérature et peuvent être classées dans les différentes catégories suivantes (Tsoumakas *et al.*, 2009) :

- Méthode d'ordonnement (*ordering*). Cette approche consiste tout d'abord à ordonner les classifieurs de base selon une fonction d'évaluation, puis à sélectionner successivement les classifieurs de base dans cet ordre fixé pour composer le sous-ensemble. L'élagage *Kappa* (*Kappa pruning*) (Margineantu *et al.*, 1997) est la première approche qui a été proposée pour l'élagage d'ensembles. Cette méthode ordonne toutes les paires de classifieurs en fonction de leur diversité, mesurée par le test statistique  $\kappa$ . Elle tente de sélectionner le sous-ensemble de classifieurs les plus divers. L'ordonnement selon l'orientation (*orientation ordering*) (Martínez-



Muñoz *et al.*, 2006) est une autre méthode efficace d'élagage basée sur l'ordonnement qui ordonne les classifieurs en fonction de la valeur de l'angle entre leur vecteur signature et le vecteur de référence. Le vecteur signature d'un classifieur  $C_i$  est un vecteur de dimension  $N$  dont les éléments prennent la valeur +1 si  $C_i(x_i) = y_i$ , et -1 si  $C_i(x_i) \neq y_i$ . Le vecteur signature de l'ensemble est la moyenne des vecteurs signatures de chaque classifieur, et le vecteur qui lui est perpendiculaire est le vecteur de référence. L'ordonnement en fonction de l'orientation privilégie le sous-ensemble qui classifie correctement les instances qui ont été mal classifiées par l'ensemble complet. La méthode d'ordonnement selon la précision (*accuracy ordering*) ordonne tous les classifieurs en fonction de leur précision de classification, puis associe les  $M$  premiers (les plus précis) classifieurs pour constituer un sous-ensemble.

- Méthode de partitionnement (*clustering*), qui utilise un algorithme de partitionnement pour trouver des groupes de classifieurs de faible diversité, puis chaque partition (*cluster*) de classifieurs est élaguée (*pruned*) séparément afin d'augmenter la diversité totale de l'ensemble. Parmi les techniques les plus connues, on peut citer, selon les différentes méthodes de partitionnement associées à l'algorithme d'élagage, l'élagage par partitionnement hiérarchique agglomératif (*hierarchical agglomerative clustering pruning*) (Giacinto *et al.*, 2000), l'élagage par  $K$  means (Lazarevic *et al.*, 2001) et le recuit déterministe (*deterministic annealing*) (Bakker *et al.*, 2003).

- Méthode d'optimisation. L'élagage d'ensembles est aussi un problème d'optimisation : *trouver le sous-ensemble de classifieurs du système de classification multiple original qui optimise une fonction d'évaluation*. Cette fonction s'appuie souvent sur des critères typiques en apprentissage automatique tels que la précision, la marge ou la diversité. Les méthodes de cette catégorie sont désignées par le processus d'optimisation utilisé dans la conception de leur modèle, telles que les méthodes de programmation semi-définies (*semi-definite programming*) (Zhang *et al.*, 2006), les algorithmes génétiques (Zhou *et al.*, 2003) et les méthodes d'escalade (*hill climbing*) (Yang *et al.*, 2005). Ces approches sont beaucoup plus coûteuses que les méthodes des catégories précédentes car elles requièrent des procédures d'optimisation non linéaire prohibitives en temps de calcul.

Le tableau 1 montre les complexités en temps de calcul et en espace mémoire des différentes méthodes d'élagage d'ensembles évoquées précédemment (Martínez-Muñoz *et al.*, 2009).  $P$  représente la taille de la population et  $E$  le nombre d'itérations dans les algorithmes génétiques. La programmation semi-définie est non seulement la méthode la plus lente mais elle est aussi dépendante de l'implémentation (Zhang *et al.*, 2006). Les techniques les plus simples sont deux méthodes de type ordonnancement. Martínez-Muñoz *et al.* (Martínez-Muñoz *et al.*, 2009) ont montré que les méthodes d'élagage par ordonnancement sont compétitives en termes de précision de classification. Nous allons proposer dans la suite une autre méthode efficace d'élagage par ordonnancement qui utilise des critères d'entropie pour sélectionner les membres du sous-ensemble.

Tableau 1. Complexité de différentes méthodes d'élagage d'ensembles

Méthode d'élagage	Complexité en temps	Complexité en mémoire
Ordonnement par précision	$O(T * N)$	$O(T * N)$
Ordonnement par orientation	$O(T * N)$	$O(T * N)$
Partitionnement	$O(T^2 * N)$	$O(T^2 * N)$
Kappa	$O(T^2 * N *  y )$	$O(T * N +  y )$
Algorithmes génétiques	$O(T * N *  y  * E * P) + O(T * E * P)$	$O(T * N) + O(T * P)$
Programmation semi-définie	$O(T^3 + T^2 * N *  y ) + O(T^3)$	$O(T^2) + O(T * N)$

#### 4. Une nouvelle méthode d'élagage d'ensemble par ordonnancement

##### 4.1. Entropie

En théorie de l'information, l'entropie (Shannon, 1948), désignée par  $H(X)$ , mesure la quantité d'information portée par une variable aléatoire  $X$ , définie par l'équation [3], où  $x_i$  représente une des instances de  $X$ ,  $I\{x_i\}$  est la quantité d'information de l'instance  $x_i$ , et  $p(x_i)$  est la fréquence des observations prenant la valeur  $x_i$ . Ainsi, l'entropie est l'espérance mathématique de toutes les quantités d'information des instances de  $X$ .

$$H(X) = E\{I(x_i)\} = -\frac{1}{|X|} \sum_{i=1}^{|X|} \log(p(x_i)) \quad [3]$$

Nous proposons une formulation originale de l'entropie  $H_m(X)$  intégrant notre nouvelle définition de la marge des méthodes d'ensemble :

$$H_m(X) = E\{I(x_i)\} = -\frac{1}{|X|} \sum_{i=1}^{|X|} \log(marge(x_i)) \quad [4]$$

Ce nouveau critère d'entropie est cohérent avec ce qui a été déclaré auparavant au sujet de la pertinence des informations qui est liée à la marge des instances. Plus la marge de l'échantillon est faible, plus la quantité d'information associée est importante, et ainsi, plus l'entropie de la variable aléatoire correspondante  $X$  est significative.

Dans la suite, le critère d'entropie introduit va être utilisé pour l'élagage d'ensembles.

#### 4.2. Élagage d'ensembles par ordonnancement selon l'entropie

Comme il a été mentionné auparavant, nous nous focalisons, pour une meilleure classification, sur les instances de plus faible marge qui sont donc plus proches des frontières des classes, et ainsi portent potentiellement des informations plus riches. En outre, les bords de classes sont les plus problématiques à classifier et ont tendance à être mal classifiés, un problème souvent rencontré en classification.

Nous proposons ici une méthode d'élagage d'ensembles par ordonnancement, qui s'appuie sur la mesure d'entropie de chaque classifieur de base de l'ensemble. Il s'agit de la nouvelle définition de l'entropie, intégrant le concept de marge, introduite dans la section précédente.

Notre méthode est basée sur la stratégie *surproduire-et-choisir* (*overproduce-and-choose*) (Partridge *et al.*, 1996) qui consiste en deux étapes : celle de surproduction et celle de choix. Dans la phase de surproduction, des classifieurs appropriés sont construits en grand nombre, ainsi les surproduisant. Dans la phase de choix, tout d'abord, les classifieurs surproduits,  $C_t$ ,  $t = 1, \dots, T$ , sont utilisés pour classifier un échantillon d'élagage (*pruning set*)  $V$  et calculer la marge, définie par l'équation [2], de chaque instance de cet échantillon. Ensuite, notre mesure d'entropie, basée sur la marge, est calculée pour chaque classifieur  $C_t$  :

$$\begin{aligned} H_m(C_t) &= E\{I(x_i)\} \\ &= -\frac{1}{N} \sum_{i=1}^N \log(\text{marge}(x_i)) \forall (x_i, y_i) \in V / C_t(x_i) = y_i \end{aligned} \quad [5]$$

où  $x_i$  représente une instance de  $V$  qui a été bien classifiée par le classifieur  $C_t$ .  $N$  est le cardinal de  $V$ .

Tous les classifieurs sont ensuite ordonnés en fonction de leur entropie. On obtient alors une liste de classifieurs triés, ayant potentiellement une fiabilité décroissante. Finalement, les  $M$  premiers classifieurs sont choisis pour constituer un ensemble de classifieurs élagué qui optimise une fonction d'évaluation sur l'échantillon d'élagage. Cette fonction intègre un ou plusieurs critères utilisés en apprentissage automatique comme la précision globale de classification, la marge ou la diversité. Des critères plus spécifiques, tels que la précision de classification de la classe mineure, souvent la plus problématique, peuvent aussi être utilisés.

Notre nouveau critère d'entropie favorise délibérément les classifieurs les plus performants dans la classification d'instances de plus faible marge. Notre méthode d'élagage d'ensemble s'apparente en quelque sorte à une stratégie de *boosting* qui se focalise aussi sur les instances les plus problématiques. Notre stratégie de sélection de sous-ensemble de classifieurs va donc aboutir à un sous-groupe de classifieurs avec une capacité potentiellement plus forte à classifier des données complexes en général, et des données en bordure de classes en particulier. En conséquence, elle va fournir des sous-ensembles de classifieurs conçus pour gérer les classes mineures de façon efficace.

### 4.3. Algorithme

Dans ce travail, nous avons utilisé la méthode de *bagging* pour créer un ensemble, et les arbres de classification et de régression (*Classification and Regression Trees (CART)*) comme classifieurs de base (Breiman *et al.*, 1984). La fonction d'évaluation utilisée pour sélectionner le meilleur sous-ensemble est la précision globale de classification. Notre algorithme consiste en les étapes suivantes :

1) Utiliser le *bagging*, à partir de la base d'apprentissage, pour générer  $T$  CART  $C_1, C_2, \dots, C_T$ . Chaque CART a été élagué (*pruned*) en changeant juste la taille minimum des nœuds terminaux de l'arbre.

2) Classifier l'échantillon d'élagage (*pruning set*) par ces CART et calculer la marge de chaque instance de cette base.

3) Calculer l'entropie  $H_m(C_t)$  basée sur la marge (définie par l'équation [5]) de chaque CART  $C_t$ , puis ordonner ces CART en fonction de leurs entropies. Rappelons que seules les instances bien classifiées sont prises en compte dans le calcul de l'entropie. L'ordre aléatoire d'origine des classifieurs de base  $C_1, C_2, \dots, C_T$  est ainsi remplacé par une séquence ordonnée  $C'_1, C'_2, \dots, C'_T$  telle que  $H_m(C'_t) > H_m(C'_{t+1}), t \leq (T - 1)$ .

4) Choisir les  $M$  premiers CART pour constituer l'ensemble élagué qui a engendré la meilleure précision globale de classification pour l'échantillon d'élagage. La taille  $M$  de l'ensemble élagué représente donc le nombre minimal de CART conduisant à la précision globale maximale de classification de l'échantillon d'élagage. Plus formellement,  $M$  est calculé de la façon suivante :

$M = |S|$  avec :

$$\begin{aligned} S &= \{C'_1, \dots, C'_M\} \\ &= \arg \max_{t, t \leq T} \text{précision}(C = \{C'_1, \dots, C'_t\}) / H_m(C'_1) > \dots > H_m(C'_t) \end{aligned} \quad [6]$$

5) Evaluer le sous-ensemble de classifieurs obtenu en classifiant la base de test.

Cet algorithme a été implémenté sur le projet R (*R project*) (R Development Core Team, 2009) et a permis l'analyse empirique décrite dans la section 5.

### 4.4. Discussion

Contrairement aux méthodes d'élagage par ordonnancement existantes, où chaque classifieur est évalué de façon indépendante, notre algorithme utilise une évaluation plus globale à travers l'entropie de marge. En effet, ce critère d'ordonnancement intègre des valeurs de marge d'instances qui sont issues d'un vote majoritaire impliquant l'ensemble complet de classifieurs. Ainsi, notre technique d'élagage n'est pas uniquement basée sur des propriétés individuelles des membres de l'ensemble (comme par exemple la précision des classifieurs de base). Elle prend en compte aussi une forme de complémentarité des classifieurs. D'autre part, la probabilité que deux classifieurs consécutifs, ordonnés selon notre critère d'entropie de marge, ne génèrent pas les mêmes erreurs est forte, parce qu'ils ont été ordonnés

selon leur capacité à classifier correctement les instances difficiles. Ainsi, cet ordonnancement, basé sur les instances de faible marge, induit plus de diversité dans la séquence de classifieurs ordonnés (constituant le sous-ensemble) que dans les autres méthodes d'ordonnancement, de même complexité, telle que la méthode d'ordonnancement selon la précision. Au contraire, cette dernière, qui est basée sur une mesure d'évaluation plus simpliste (précision individuelle), a une forte probabilité que des classifieurs ordonnés consécutifs commettent les mêmes erreurs, ainsi induisant moins de diversité, puisqu'ils s'appuient sur des instances de plus forte marge. Les classifieurs dont les erreurs ne sont pas corrélées sont dits complémentaires. Notre méthode aura donc tendance à sélectionner des classifieurs complémentaires et à maintenir une grande diversité entre les membres sélectionnés du sous-ensemble retenu. Elle est basée sur un critère d'ordonnancement qui intègre de la diversité, mais pas exclusivement comme l'élagage *Kappa*, et de la précision mais pas exclusivement comme l'élagage par ordonnancement selon la précision. Il est bien connu que ni la précision des classifieurs de base, ni la diversité ne suffisent à elles seules pour identifier des ensembles performants. Il est nécessaire de prendre en compte la complémentarité des classifieurs (Martínez-Muñoz *et al.*, 2009). Par conséquent, notre approche a plus de potentiel pour construire des ensembles efficaces que d'autres méthodes simples basées seulement sur des propriétés individuelles des classifieurs de l'ensemble.

## 5. Résultats expérimentaux

### 5.1. Jeux de données

Tableau 2. Jeux de données

Jeux de données	Apprentissage	Élagage	Test	Attributs	Classes
Connect-4	2000	2000	2000	42	3
Glass	72	71	71	9	6
Kr-vs-kp	1065	1065	1065	36	2
Letter	2000	2000	2000	16	26
Optdigits	1000	1000	1000	64	10
Pendigit	2000	2000	2000	16	10
Pima	256	256	256	8	2
Tic-tac	310	310	310	9	2
Waveform	1000	1000	1000	21	3
Wine quality-red	533	533	533	11	6
Image aérienne	8000	94265	93395	3	4

Pour évaluer notre schéma d'élagage d'ensembles par ordonnancement de classifieurs en fonction de l'entropie, nous avons effectué des expérimentations sur 10 jeux de données représentatifs de la base UCI (Asuncion *et al.*, 2007), et sur une image aérienne d'un site urbain montrée sur la figure 2. Cette image présente quatre classes d'intérêt : *sol artificiel*, *sol naturel*, *bâtiment* et *végétation* qui sont représentées par 3 attributs : les composantes RGB de l'image. Chaque jeu de données

a été divisé en trois parties : l'échantillon d'apprentissage, l'échantillon d'élagage (*pruning set*) et l'échantillon de test, comme le montre le tableau 2. Pour tous les jeux de données, la taille de l'ensemble de classifieurs initial a été fixée à 500.



Figure 2. Image aérienne d'un site urbain

## 5.2. Analyse comparative de la performance d'élagage d'ensembles

Dans cette analyse empirique, notre méthode d'élagage d'ensembles est confrontée à l'approche naïve d'élagage aléatoire de classifieurs de base, ainsi qu'à un autre algorithme d'élagage d'ensembles par ordonnancement, à savoir la technique d'élagage par ordonnancement selon la précision. Tous les résultats présentés dans la suite sont issus d'un moyennage sur 10 passes de calcul.

### 5.2.1. Performance de classification globale

Le tableau 3 présente la moyenne et l'écart type de la précision de classification obtenue sur la base de test par le sous-ensemble de classifieurs qui a engendré la précision de classification maximale sur l'échantillon d'élagage pour les deux méthodes d'élagage d'ensembles par ordonnancement. La dernière colonne de ce tableau exhibe la précision de classification obtenue sans procédure d'élagage, autrement dit, elle est produite par l'ensemble complet des classifieurs de base. Ce tableau montre clairement que notre méthode a une meilleure précision de classification que la méthode d'élagage d'ensembles par ordonnancement selon la précision ainsi que celle produite par l'ensemble complet de classifieurs. Il est à noter que notre technique augmente la précision de presque 10 % par rapport à l'utilisation brute de tous les classifieurs (pas d'élagage) sur le jeu de données *Tic-tac*.

Nous avons également utilisé le test statistique *Kappa de Cohen* (Cohen, 1960) comme critère d'évaluation de performance des algorithmes d'élagage d'ensembles. Ce test mesure l'accord (ou désaccord) entre observateurs lors d'un jugement

qualitatif en catégories. Une partie de l'accord (ou désaccord) pouvant être due au hasard, ce test tente de corriger le degré d'accord en lui déduisant la partie due au hasard. Le test du Kappa se calcule de la manière suivante :

$$\kappa = \frac{P_o - P_c}{1 - P_c} \quad [7]$$

où  $P_o$  est la probabilité d'accord relatif entre observateurs et  $P_c$  la probabilité d'un accord aléatoire. Les valeurs du coefficient  $\kappa$  varient de  $-1$  (désaccord total entre observateurs) à  $1$  (observateurs complètement en accord) en passant par  $0$  (accord dû seulement au hasard).

Tableau 3. Précision de classification (moyenne  $\pm$  écart type) du sous-ensemble sélectionné sur la base de test

Jeu de données	Ordonnancement par entropie (%)	Ordonnancement par précision (%)	Sans élagage (%)
Connect-4	73,94 $\pm$ 0,26	72,95 $\pm$ 0,57	72,65 $\pm$ 0,25
Glass	68,73 $\pm$ 2,95	67,04 $\pm$ 2,67	61,97 $\pm$ 0,13
Kr-vs-kp	99,00 $\pm$ 0,26	98,96 $\pm$ 0,26	97,06 $\pm$ 0,09
Letter	71,16 $\pm$ 0,64	68,46 $\pm$ 0,44	68,18 $\pm$ 0,33
Optdigits	95,12 $\pm$ 0,51	93,44 $\pm$ 0,42	93,59 $\pm$ 0,25
Pendigit	94,85 $\pm$ 0,17	92,96 $\pm$ 0,58	92,50 $\pm$ 0,16
Pima	72,58 $\pm$ 1,66	70,55 $\pm$ 1,72	68,55 $\pm$ 0,53
Tic-tac	84,94 $\pm$ 1,18	83,06 $\pm$ 1,22	76,29 $\pm$ 0,49
Waveform	81,32 $\pm$ 0,97	79,41 $\pm$ 0,55	80,35 $\pm$ 0,41
Wine quality-red	62,05 $\pm$ 1,28	61,83 $\pm$ 1,60	60,78 $\pm$ 0,43
Image aérienne	81,63 $\pm$ 0,08	81,47 $\pm$ 0,06	81,43 $\pm$ 0,03

Le tableau 4, structuré comme le tableau précédent, confirme la supériorité de notre méthode par rapport à la méthode d'élagage d'ensembles par ordonnancement selon la précision ainsi que vis-à-vis de la version sans élagage.

La figure 3 montre les courbes de précision de classification, en fonction de la taille du sous-ensemble sélectionné, pour les méthodes d'élagage par ordonnancement selon l'entropie, par ordonnancement selon la précision, et par ordonnancement aléatoire, sur le jeu de données *Pendigit*. La précision de l'agrégation aléatoire augmente lentement et de façon monotone en fonction du nombre de classifieurs impliqués jusqu'à atteindre un niveau de précision stable pour un ensemble *saturé*. Notre technique a non seulement réduit la complexité d'un facteur 5 (25 classifieurs requis au lieu de 131), mais aussi augmenté la précision de classification de 2,6 % par rapport à la technique d'agrégation aléatoire. Les courbes de précision des deux méthodes d'agrégation par ordonnancement (entropie et précision) atteignent un optimum de précision, puis décroissent lentement et de façon monotone jusqu'à atteindre la précision de l'ensemble complet. Néanmoins, notre méthode d'élagage par ordonnancement en fonction de l'entropie engendre un pic de

précision nettement plus haut que celui obtenu par la méthode d'élagage par ordonnancement en fonction de la précision. L'évolution de ces deux courbes est conforme aux résultats d'élagage d'ensemble par ordonnancement présentés dans (Martínez-Muñoz *et al.*, 2009). L'optimum de précision atteint correspond au sous-ensemble de classificateurs dont la précision est meilleure que celle de l'ensemble initial, sans élagage. Comme la combinaison des classificateurs de base s'effectue par vote majoritaire, l'influence de ce sous-ensemble de classificateurs les plus performants s'estompe quand le nombre de classificateurs considérés augmente, d'où la décroissance de ces courbes de précision.

Tableau 4. Test du Kappa (moyenne  $\pm$  écart type) du sous-ensemble sélectionné sur la base de test

Jeu de données	Ordonnancement par entropie (%)	Ordonnancement par précision (%)	Sans élagage (%)
Connect-4	36,28 $\pm$ 1,44	33,63 $\pm$ 2,10	31,05 $\pm$ 0,29
Glass	61,10 $\pm$ 3,82	62,11 $\pm$ 3,45	59,28 $\pm$ 1,58
Kr-vs-kp	98,83 $\pm$ 0,19	98,73 $\pm$ 0,44	98,52 $\pm$ 0,07
Letter	68,31 $\pm$ 0,33	66,93 $\pm$ 0,45	66,25 $\pm$ 0,36
Optdigits	94,83 $\pm$ 0,41	93,82 $\pm$ 0,61	94,37 $\pm$ 0,15
Pendigit	93,19 $\pm$ 0,51	91,92 $\pm$ 0,16	91,72 $\pm$ 0,14
Pima	43,88 $\pm$ 4,10	42,23 $\pm$ 4,82	44,76 $\pm$ 0,80
Tic-tac	57,17 $\pm$ 4,21	52,26 $\pm$ 4,77	44,27 $\pm$ 2,10
Waveform	72,25 $\pm$ 1,11	68,99 $\pm$ 0,88	70,82 $\pm$ 0,64
Wine quality-red	40,70 $\pm$ 1,86	39,73 $\pm$ 2,66	38,88 $\pm$ 0,61
Image aérienne	66,30 $\pm$ 0,05	66,00 $\pm$ 0,04	65,38 $\pm$ 0,02

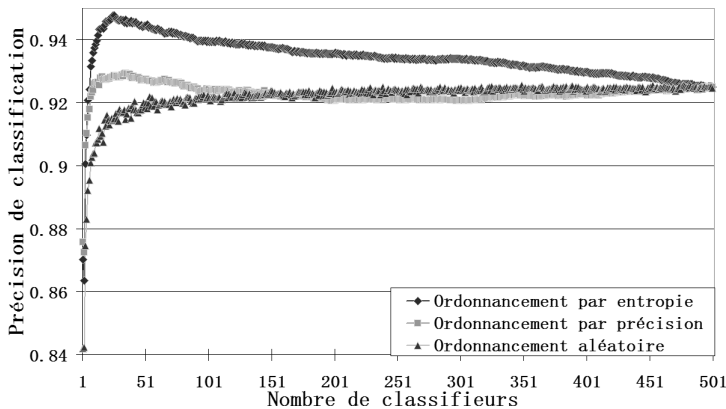


Figure 3. Précision de classification moyenne en fonction du nombre de classificateurs des méthodes d'élagage aléatoire, par ordonnancement selon la précision, et par ordonnancement selon l'entropie, sur le jeu de données Pendigit



La figure 4 montre la validation de notre algorithme d'élagage d'ensemble sur la base de test d'une image aérienne (voir figure 2). Notre technique a réduit la complexité d'un facteur 4 (22 classifieurs requis au lieu de 93) par rapport à la méthode d'élagage aléatoire d'ensemble.

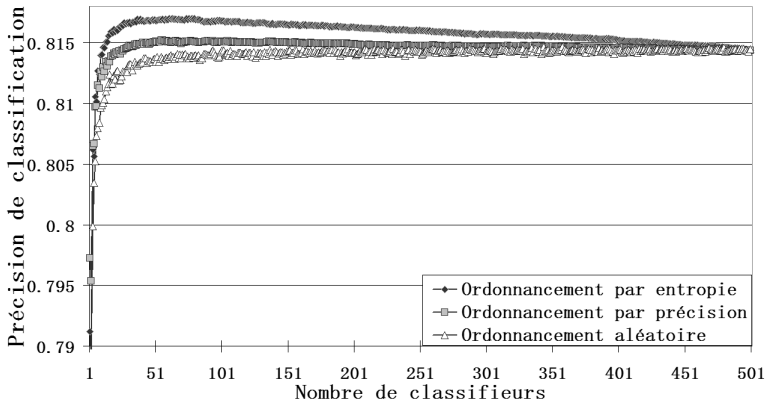


Figure 4. Précision de classification moyenne des méthodes d'agrégation par ordonnancement selon l'entropie, par ordonnancement selon la précision et par ordonnancement aléatoire sur la base de test de l'image aérienne

### 5.2.2. Performance de classification par classe

La figure 5 montre les courbes de précision de classification, en fonction du nombre de classifieurs, des deux méthodes d'élagage par ordonnancement pour deux des classes de *Pendigit*. Il apparaît que notre méthode augmente la précision de la classe 3 de façon significative, contrairement à la méthode concurrente. Cela est dû au fait que notre approche privilégie les classifieurs qui classifient efficacement les instances de faible marge. En effet, la classe 3 a deux fois plus d'instances de faible marge que la classe 0.

### 5.2.3. Complexité

Notre technique est de complexité moindre, comme le montre la figure 6 qui exhibe, pour tous les jeux de données, le nombre minimum de classifieurs de base constituant le sous-ensemble sélectionné, par les deux méthodes d'élagage par ordonnancement, sur l'échantillon d'élagage.

La figure 7 exhibe la différence,  $\Delta M = |M_e - M_t|$ , entre le nombre de classifieurs requis,  $M_e$ , sur l'échantillon d'élagage, et le nombre de classifieurs requis,  $M_t$ , sur l'échantillon de test, par les deux méthodes d'élagage par ordonnancement, pour tous les jeux de données. Ce résultat met en évidence une meilleure stabilité de notre méthode en termes de complexité, autrement dit, de la taille minimale du sous-ensemble de classifieurs choisis. La méthode d'élagage par ordonnancement selon la précision s'avère particulièrement incohérente, et donc imprévisible, pour les jeux de test *Opdigits* et *Waveform*. Cela pourrait provenir du

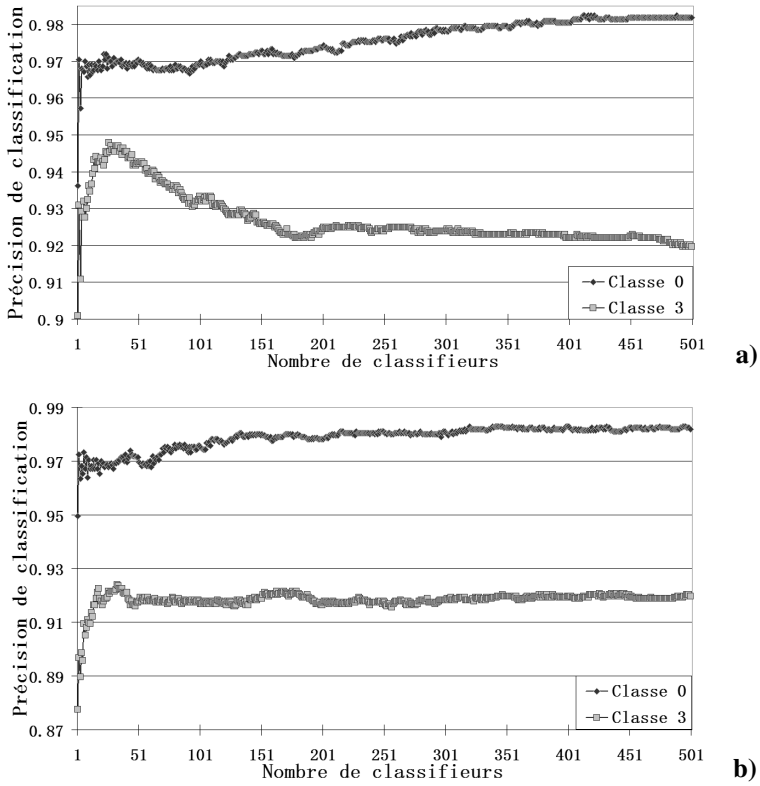


Figure 5. Précision de classification moyenne des méthodes d'agrégation par ordonnancement selon l'entropie a) et par ordonnancement selon la précision b) pour deux des classes de Pendigit sur l'ensemble de test

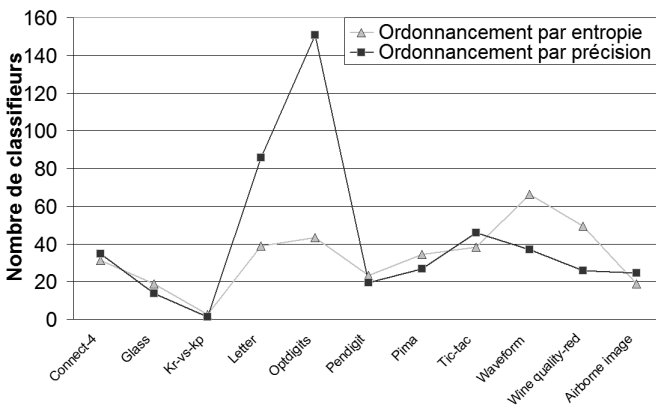


Figure 6. Nombre de classifieurs de base sélectionnés par les deux méthodes d'élagage par ordonnancement, sur l'échantillon d'élagage, pour tous les jeux de données

fait que cette méthode est confrontée au problème de surajustement (*overfitting*) des modèles (Dietterich, 1995) de l'échantillon d'élagage. En revanche, notre méthode engendre la précision maximale avec des tailles de sous-ensemble du même ordre sur les échantillons d'élagage et de test. Le fait que l'on puisse prédire de façon relativement précise, *via* l'échantillon de validation du modèle (ou échantillon d'élagage), le nombre minimum de classifieurs de base à utiliser pour classifier l'échantillon de test de façon optimale explique en partie pourquoi notre méthode est plus performante aussi en termes de précision de classification, comme le montre le tableau 3.

#### 5.2.4. Distributions de marges

La figure 8 montre les distributions de marges (des instances bien classifiées) cumulées obtenues par le sous-ensemble de classifieurs sélectionné pour les méthodes d'élagage par ordonnancement selon l'entropie, par ordonnancement selon la précision, et sans élagage, sur l'échantillon d'élagage de *Pendigit*. Notre méthode exhibe une *meilleure* dispersion des valeurs de marge qui se traduit par un taux plus important des marges moyennes. La méthode d'ordonnancement selon la précision montre une progression plus significative sur les fortes marges (au-delà de 0,9). Néanmoins, cette nouvelle dispersion des marges par notre algorithme, qui consiste à augmenter les valeurs de marge faible à moyenne, a un impact plus favorable sur la qualité de classification que la nouvelle dispersion associée à la méthode basée sur la précision, qui, elle, favorise les très fortes marges. Augmenter les marges des exemples d'apprentissage est corrélée avec la réduction de l'erreur de prédiction (base de test) (Schapire *et al.*, 1998). Il est clair qu'augmenter les valeurs des marges les plus faibles est plus bénéfique qu'augmenter les marges les plus fortes (bénéficiant déjà d'une mesure de confiance importante), comme en témoignent les résultats de performance de classification présentés précédemment, qui sont en faveur de notre approche.

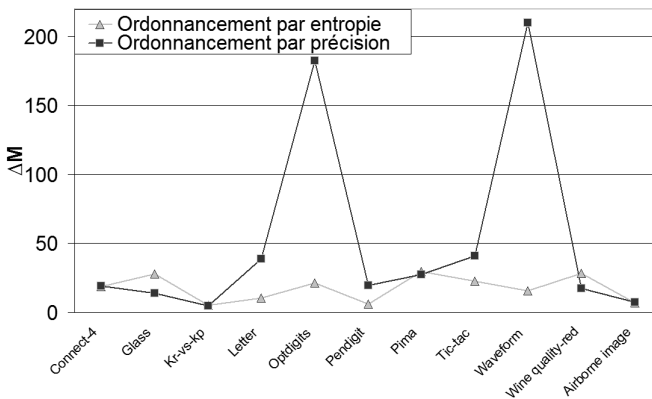


Figure 7. Différence,  $\Delta M = |M_e - M_t|$ , entre le nombre de classifieurs,  $M_e$ , requis sur l'échantillon d'élagage, et le nombre de classifieurs,  $M_t$ , requis sur l'échantillon de test, par les deux méthodes d'élagage par ordonnancement, pour tous les jeux de données

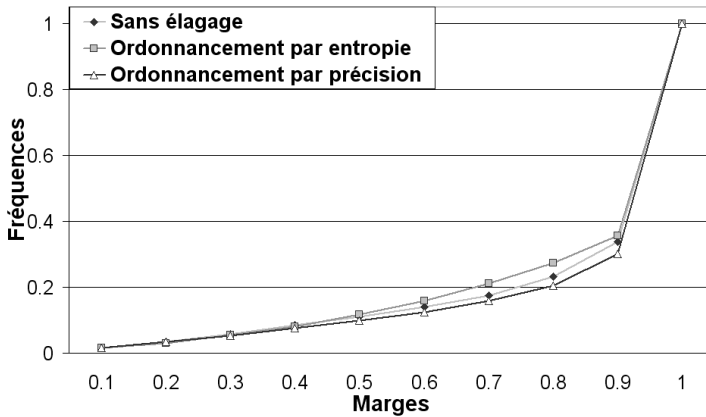


Figure 8. Distributions de marges (des instances bien classifiées) cumulées obtenues par le sous-ensemble sélectionné pour les méthodes d'élagage par ordonnancement selon l'entropie, par ordonnancement selon la précision, et sans élagage, sur l'échantillon d'élagage de Pendigit

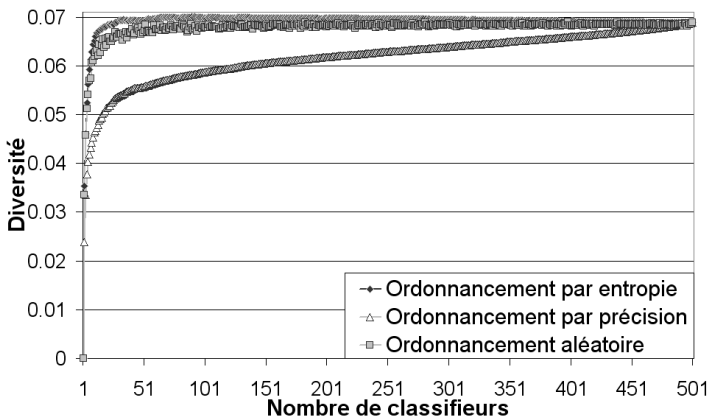


Figure 9. Diversité des sous-ensembles sélectionnés par les méthodes d'élagage par ordonnancement selon l'entropie, par ordonnancement selon la précision, et aléatoire, sur l'échantillon de test de Pendigit

### 5.2.5. Diversité

Pour évaluer la diversité des sous-ensembles de classifieurs engendrés par les méthodes d'élagage par ordonnancement selon l'entropie, par ordonnancement selon la précision, et aléatoire, nous utilisons l'une des mesures de diversité les plus populaires : la mesure *KW* (Kohavi *et al.*, 1996). La figure 9 montre les courbes de diversité en fonction du nombre de classifieurs ordonnés inclus dans le sous-ensemble par les 3 méthodes d'élagage sur le jeu de données *Pendigit*. On peut noter

que notre méthode exhibe plus de diversité et converge plus rapidement que les autres méthodes. Même si l'augmentation de diversité par rapport à la méthode d'ordonnancement aléatoire est légère, elle est significative comparée à l'ordonnancement selon la précision. En outre, il est bien connu que l'aléatoire engendre une grande diversité mais sans pour autant induire la meilleure précision. Ces résultats confirment que notre méthode a tendance à produire des classifieurs plus *divers* et donc plus complémentaires comme souligné dans le paragraphe 4.4.

## 6. Conclusion

Nous avons présenté un algorithme original d'élagage d'ensembles par ordonnancement, en fonction de l'entropie, exploitant une nouvelle définition de la marge des méthodes d'ensembles. Notre stratégie d'élagage considère les instances de plus faible marge comme étant les plus pertinentes pour la construction de classifieurs fiables. Notre méthode réduit de façon très significative la complexité des méthodes d'ensembles. Elle assure aussi une meilleure précision de classification que la version sans élagage, en particulier dans le cas de classes difficiles. Cette approche s'est avérée aussi plus performante que la méthode d'élagage par ordonnancement en fonction de la précision. En outre, notre algorithme a du potentiel pour classifier des jeux de données déséquilibrés puisqu'il exploite des instances de faible marge. Actuellement, nous analysons la capacité de notre algorithme à gérer ce problème délicat. Une analyse plus profonde de la nouvelle définition de marge d'ensemble que nous proposons et de son intérêt vis-à-vis de la formulation classique est aussi en cours. Une des perspectives, à court terme, de ce travail serait d'élaborer une stratégie efficace pour sélectionner le meilleur sous-ensemble de classifieurs par ordonnancement selon l'entropie. En effet, le nombre optimal de classifieurs de base, entraînant la meilleure précision, peut varier de manière significative d'une classe à l'autre ; et une sélection basée sur la précision globale de classification est l'approche la plus simple à considérer. Nous l'avons utilisée juste pour mettre en évidence l'efficacité de notre méthode d'élagage par rapport à d'autres méthodes existantes mais certainement pas pour atteindre la meilleure performance de classification. En outre, la nouvelle marge que nous avons proposée est potentiellement plus robuste au bruit que la marge d'ensemble classique puisqu'elle n'est pas affectée par des erreurs d'étiquetage de certaines instances. Ainsi, une autre perspective de ce travail serait d'étudier la performance des ensembles élagués, autrement dit après une procédure de *pruning*, dans la résolution des problèmes de classification entachés de bruit d'étiquetage.

## Remerciements

*Les auteurs souhaiteraient remercier le laboratoire MATIS de l'institut IGN pour la mise à disposition de l'image aérienne et la vérité terrain associée.*

## Bibliographie

- Asuncion A., Newman D. (2007), UCI Machine Learning Repository.
- Bakker B., Heskes T. (2003), Clustering ensembles of neural network models, *Neural Networks*, vol. 16, n°2, p. 261-269.
- Banfield R. E., Hall L. O., Bowyer K. W., Kegelmeyer W. P. (2005), Ensemble diversity measures and their application to thinning, *Information Fusion*, vol. 6, n°1, p. 49-62.
- Bartlett P. J., Schölkopf B., Schuurmans D., Smola, A. J. (eds) (2000), *Advances in Large Margin Classifiers*, Neural Information Processing, 1 edn, The MIT Press.
- Bauer E., Kohavi R. (1999), An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants, *Machine Learning*, vol. 36, n°1, p. 105-139.
- Breiman L. (1996), Bagging predictors, *Machine Learning*, vol. 24, n°2, p. 123-140.
- Breiman L., Friedman J., Olshen R., Stone C. (1984), *Classification and Regression Trees*, Publisher : Wadsworth.
- Brown G., Wyatt J., Harris R., Yao X. (2005), Diversity Creation Methods: A Survey and Categorisation, *Journal of Information Fusion*, vol. 6, n°1, p. 5-20.
- Chehata N., Guo L., Mallet C. (2009), Contribution of Airborne Full-Waveform Lidar and Image Data for Urban Scene Classification, *ICIP'2009, 16th IEEE International Conference on Image Processing*, p. 1667-1672.
- Cohen J. (1960), A Coefficient of Agreement for Nominal Scales, *Educational and Psychological Measurement*, vol. 20, n°1, p. 37-46
- Cutler D. R., Edwards T. C., Beard K. H., Cutler A., Hess K. T., Gibson J., Lawler J. J. (2007), Random forests for classification in ecology, *Ecology*, vol. 88, n°11, p. 2783-2792.
- Dietterich T. G. (1995), Overfitting and under-computing in machine learning, *Computing Surveys*, vol. 27, n°3, p. 326-327.
- Dietterich T. G. (1997), Machine-Learning Research: Four Current Directions, *AI Magazine*, vol. 18, p. 97-136.
- Dietterich T. G. (2000), Ensemble Methods in Machine Learning, *1st International Workshop on Multiple Classifier Systems*, p. 1-15.
- Efron B., Tibshirani R. (1994), *An Introduction to the Bootstrap*, Chapman & Hall/CRC.
- Freund Y., Schapire R. E. (1997), A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, vol. 55, n°1, p. 119-139.
- Garey M., Johnson D. (1979), *Computers and Intractability : A Guide to the Theory of NP-Completeness*, Publisher : W. H. Freeman.
- Giacinto G., Roli F., Fumera G. (2000), Design of Effective Multiple Classifier Systems by Clustering of Classifiers, *ICPR'2000, 15th International Conference on Pattern Recognition*, p. 160-163.
- Gislason P., Benediktsson J., Sveinsson J. (2006), Random Forests for land cover classification, *Pattern Recognition Letters*, vol. 27, n°4, p. 294-300.

- Guo L., Boukir S., Chehata N. (2010a), Support vectors selection for supervised learning using an ensemble approach, *ICPR'2010, 20th IAPR International Conference on Pattern Recognition*, p. 37-40.
- Guo L., Chehata N., Boukir S. (2010b), A two-pass random forests classification of airborne lidar and image data on urban scenes, *ICIP'2010, 17th IEEE International Conference on Image Processing*, p. 1369-1372.
- Ho T. K. (1998), The Random Subspace Method for Constructing Decision Forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, n°8, p. 832-844.
- Kohavi R., Wolpert D. H. (1996), Bias Plus Variance Decomposition for Zero-One Loss Functions, *13th International Conference of Machine Learning, ICML'96*, p. 275-283.
- Kuncheva L. (2004), *Combining Pattern Classifiers : Methods and Algorithms*, Wiley-Interscience. ISBN : 0471210781.
- Kuncheva L. I. (2002), Switching Between Selection and Fusion in Combining Classifiers: An Experiment, *IEEE Transactions on Systems, Man, and Cybernetics, Part B : Cybernetics*, vol. 32, n°2, p. 146-156.
- Kuncheva L. I., Whitaker C. J. (2003), Measures of diversity in classifier ensembles and Their Relationship with the Ensemble Accuracy, *Machine Learning*, vol. 51, p. 181-207.
- Lazarevic A., Obradovic Z. (2001), Effective pruning of neural network classifiers, *IEEE/INNS International Conference on Neural Networks*, p. 796-801.
- Mangiameli P., West D., Rampal R. (2004), Model selection for medical diagnosis decision support systems, *Decision Support Systems*, vol. 36, n°3, p. 247-259.
- Margineantu D. D., Dietterich T. G. (1997), Pruning Adaptive Boosting, *ICML'1997, 14th International Conference on Machine Learning*, p. 211-218.
- Martínez-Muñoz G., Hernández-Lobato D., Suárez A. (2009), An Analysis of Ensemble Pruning Techniques Based on Ordered Aggregation, *IEEE Transactions in Pattern Analysis and Machine Intelligence*, vol. 31, n°2, p. 245-259.
- Martínez-Muñoz G., Suárez A. (2006), Pruning in ordered bagging ensembles, *ICML'2006, 23rd International Conference on Machine Learning*, p. 609-616.
- Pang H., Zhao H. (2008), Building pathway clusters from Random Forests classification using class votes, *BMC Bioinformatics*.
- Partridge D., Yates W. B. (1996), Engineering multiversion neural-net systems, *Neural Computation*, vol. 8, n°4, p. 869-893.
- R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. 2009, ISBN 3-900051-07-0.
- Schapire R. E., Freund Y., Bartlett P., Lee W. S. (1998), Boosting the margin: a new explanation for the effectiveness of voting methods, *The Annals of Statistics*, vol. 26, n°5, p. 1651-1686.
- Shannon C. (1948), A mathematical theory of communication, *Bell System Technical Journal*, vol. 27, n°3, p. 379-423.
- Tao D., Tang X., Li X., Wu X. (2006), Asymmetric Bagging and Random Subspace for Support Vector Machines-Based Relevance Feedback in Image Retrieval, *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, n°7, p. 1088-1099.

- Tsoumakas G., Partalas I., Vlahavas I. (2009), *Applications of Supervised and Unsupervised Ensemble Methods*, vol. 245, Springer, chapter An Ensemble Pruning Primer, p. 1-13.
- Yang Y., Korb K., Ting K. M., Webb G. I. (2005), Ensemble Selection for SuperParent-One-Dependence Estimators, *AI 2005: Advances in Artificial Intelligence*, p. 102-112.
- Zhang Y., Burer S., Street W. N. (2006), Ensemble Pruning Via Semi-definite Programming, *Journal of Machine Learning Research*, vol. 7, p. 1315-1338.
- Zhou Z.-H., Tang W. (2003), Selective ensemble of decision trees, *9th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, p. 476-483.
- Zhou Z.-H., Wu J., Tang W. (2002), Ensembling neural networks: Many could be better than all, *Artificial Intelligence*, vol. 137, n°1-2, p. 239-263.

Article reçu le 18/06/2010

Accepté le 27/01/2011



**Samia Boukir** a obtenu le DEA de Robotique de Paris 6 en 1990 et le Doctorat en Vision-Robotique de l'INRIA de Rennes en 1993. Elle est Professeur en Informatique à l'Institut EGID – université de Bordeaux 3, depuis 2005. Elle est membre du Laboratoire Géoressources et Environnement de l'université de Bordeaux 3, depuis 2005. Ses activités de recherche concernent principalement le traitement et l'analyse de séquences spatio-temporelles (images et signaux). Elle s'intéresse tout particulièrement à l'appariement et au suivi de primitives bidimensionnelles dans une séquence d'images ainsi qu'à l'analyse du mouvement. Depuis 2008, elle travaille sur des problèmes de classification avec des applications en télédétection, notamment la cartographie de l'occupation du sol et la détection de changements environnementaux.



**Li Guo** a obtenu la Maîtrise d'Informatique de l'université du Nord Ouest de Xi'an (Chine) en 2006 et le Master d'Ingénierie de l'Image de l'Institut EGID – université de Bordeaux 3, en 2007. Il est Doctorant au Laboratoire Géoressources et Environnement de l'université de Bordeaux 3 depuis décembre 2007. Ses activités de recherche concernent principalement la reconnaissance de formes et l'apprentissage automatique, notamment les méthodes d'ensembles de classifieurs.