

Sélection de variables et agrégation d'opinions

Gaëlle Legrand et Nicolas Nicoloyannis*

*Laboratoire ERIC
Université Lumière Lyon 2
Batiment L
5 av. Pierre Mendès-France
69 676 BRON cedex FRANCE
glegrand@eric.univ-lyon2.fr ; nicolas.nicoloyannis@univ-lyon2.fr

Résumé. La taille des bases de données étant de plus en plus importante, le processus de sélection de variables devient essentiel. Nous proposons une méthode de sélection, pour les variables qualitatives, basée sur l'agrégation d'opinion. Le résultat, sous forme d'un préordre de variables, est fourni par l'agrégation des résultats obtenus par plusieurs méthodes myopes de sélection de variables.

1 Introduction

La taille des bases de données étant de plus en plus importante, l'amélioration de la qualité de représentation des données est devenue un problème majeur de l'extraction des connaissances à partir des données. L'une des difficultés principales liée à la représentation des données est la dimension des données.

Le problème de la dimension des données concerne le nombre de variables descriptives caractérisant chacun des individus. Parmi ces variables, certaines peuvent être non pertinentes, inutiles et/ou redondantes. Donc, si l'on désire extraire de l'information utile et compréhensible à partir de nos données, il convient en premier lieu de retirer les parties non pertinentes.

La sélection de variables permet de résoudre ce problème. C'est un processus choisissant un sous-ensemble optimal de variables selon un critère particulier. Il permet l'élimination de variables inutiles, non pertinentes et redondantes ainsi que l'élimination du bruit généré par certaines variables. Le processus d'apprentissage est accéléré et la précision prédictive des algorithmes d'induction peut être améliorée.

Il existe deux familles d'algorithmes de sélection de variables : les méthodes "enveloppe" [John *et al.*, 1994] et les méthodes "filtre" [Kira et Rendell, 1992a]. La différence fondamentale entre ces deux familles réside dans le fait que la première est liée à l'algorithme d'induction utilisée alors que la seconde est totalement indépendante.

1.1 Approches Enveloppe

Les méthodes de type enveloppe prennent en compte l'influence du sous-ensemble de variables sélectionné sur les performances de l'algorithme d'induction. Elles utilisent l'algorithme d'apprentissage comme fonction d'évaluation pour tester les différents sous-ensembles de variables générés. Cependant, leur coût calculatoire est bien souvent trop important.

1.2 Approches Filtre

Les approches filtre sont de 5 types :

- Les méthodes exhaustives testent tous les sous-ensembles possibles de P variables parmi M variables existantes. Nous pouvons citer les algorithmes MDLM [Sheinvald *et al.*, 1990] ou FOCUS [Almuallim et G., 1992]. La complexité de FOCUS est de l'ordre de $O(N^M)$, avec N le nombre d'individus. Ces algorithmes sont impossibles à appliquer du fait de leur coût calculatoire trop élevé.
- Les méthodes heuristiques sont très nombreuses. La méthode la plus connue est RELIEF [Kira et Rendell, 1992b] dont la complexité est de l'ordre de $O(INM)$ avec I le nombre d'itérations effectuées et fixées par l'utilisateur. Il existe également des méthodes du type Branch and bound telles que ABB [Liu *et al.*, 1998]. Ces méthodes requièrent plusieurs accès à la base de données.
- Les méthodes probabilistes sont représentées par LVF [Liu et Setiono, 1996]. Sa complexité est de l'ordre de $O(INM)$ avec I le nombre d'itérations effectuées et fixées par l'utilisateur. Cependant, du fait de sa caractéristique probabiliste, le nombre de variables sélectionnées tend vers la moitié du nombre de variables initiales. Comme les méthodes précédentes, ces méthodes requièrent plusieurs accès à la base de données.
- Les méthodes de sélection en un seul parcours de base sont des processus itératifs qui, comme leur nom l'indique, ne nécessitent qu'un seul scan de la base étudiée. Le processus de sélection s'effectue de la manière suivante : lors de la première étape, la variable la plus corrélée avec la variable endogène est sélectionnée. Lors de la deuxième étape, la variable qui est la plus partiellement corrélée avec la variable endogène est sélectionnée et ainsi de suite. Afin de n'avoir qu'un seul scan de base, les mesures rapides de corrélation sont utilisées (coefficient de corrélation linéaire de Pearson, coefficient de corrélation de rangs de Kendall, etc). Ce type de méthode est représenté par MIFS [Battiti, 1994], CFS [Hall, 2000], et la méthode proposée par Lallich et Rakotomalala, [Lallich et Rakotomalala, 2000]. Ces méthodes, qui sont les plus rapides et qui sont relativement efficaces, paraissent les plus intéressantes.
- Les méthodes myopes ou pas à pas utilisent des critères de sélection myopes pour sélectionner les variables. Ces méthodes ne prennent pas en compte l'interaction entre variables et permettent de classer les variables en fonction de leur pouvoir discriminant. Ce type de méthodes est efficace et très rapide en particulier sur des problèmes comportant à la fois beaucoup de variables et d'individus. La complexité de ce type de méthodes est de l'ordre de $O(N \log N)$.

2 Point de départ

Nous sommes partis du constat suivant : les méthodes pas à pas utilisant des critères de sélection myopes tels que l'entropie de Shannon sont rapides, peu coûteuses et présentent des résultats plutôt encourageants. Il existe 4 catégories de critères permettant de mesurer différentes caractéristiques des variables :

- Les critères d'information : c'est la quantité d'information apportée par une va-

riable sur la variable à prédire. La variable, ayant le gain d'information le plus élevé, sera préférée aux autres variables : l'entropie de Shannon [Shannon, 1948], le ratio du gain [Quinlan, 1986], le gain normalisé [Jun *et al.*, 1997].

- Les critères de distance : ces mesures s'intéressent au pouvoir discriminant d'une variable. Elles évaluent la séparabilité des classes en se basant sur les distributions de probabilités des classes : la distance de Mantaras [De Mantaras, 1991], le critère de Gini [Breiman *et al.*, 1984].
- Les critères d'indépendance regroupent toutes les mesures de corrélation ou d'association. Elles permettent de calculer le degré avec lequel une variable exogène est associée à une variable endogène. Le calcul de l'écart à l'indépendance de deux variables d'un tableau de contingence est effectué : le chi2, le critère de Tschuprow [Hart, 1984] et [Mingers, 1987], le coefficient de Cramer.
- Les critères de consistance : ils recherchent l'ensemble de variables le plus petit qui satisfait un pourcentage d'inconsistance minimum défini par l'utilisateur. Deux objets sont inconsistants si leurs modalités sont identiques et s'ils appartiennent à deux classes différentes. Ces mesures permettent de détecter les variables redondantes : le τ de Zhou [Zhou et Dillon, 1991].

Cependant, l'utilisation d'une méthode myope génère trois problèmes :

1. Le choix du critère est délicat : quel critère est le plus efficace?
2. La forme du résultat (une liste de variables triées) ne nous permet pas de déterminer le sous-ensemble optimal de variables.
3. Ce type de méthodes ne prend pas en compte l'interaction existante entre les variables exogènes. Nous verrons, lors des expérimentations, que cela n'a pas d'impact sur la qualité des résultats.

La méthode que nous proposons permet de résoudre les deux premiers problèmes soulevés de la manière suivante :

1. Il n'existe pas de critère meilleur ou plus efficace que les autres. Chaque critère met en avant certaines qualités spécifiques à chaque variable. Il semble intéressant d'obtenir un résultat tenant compte de l'avis de plusieurs critères différents. Afin d'obtenir ce type de résultats, nous utilisons une méthode d'agrégation d'opinions.
2. L'obtention d'une liste triée de variables limite l'intérêt de la sélection de variables. En effet, comment peut-on déterminer la taille optimale du sous-ensemble? Lorsque l'on est en présence d'une liste triée de variables, l'une des méthodes qui semble efficace pour obtenir un sous-ensemble optimal de variables est d'utiliser une approche enveloppe qui ajoute ou ôte itérativement les éléments de la liste triée. À chaque itération, la méthode d'apprentissage est appliquée pour tester si l'ajout ou la suppression d'une variable entraîne une amélioration du taux d'apprentissage. Toutefois, ce processus est bien trop coûteux et long pour être appliqué. Pour cette raison, nous paramétrons la méthode d'agrégation utilisée pour qu'elle nous fournisse non pas un ordre sur les variables mais un préordre total. Aussi, nous n'ajouterons pas les variables une par une mais par sous-ensemble de variables.

3 Présentation de la méthode

La méthode de sélection proposée est une méthode hybride à l'intersection des approches filtre et enveloppe. Elle est de type Forward Selection et agrège les classements des variables obtenus à l'aide de plusieurs critères de sélection myopes. Elle ne traite que les variables qualitatives. Lorsque des variables quantitatives sont présentes, elles sont discrétisées de manière supervisée à l'aide de la méthode Fusinter [Zighed *et al.*, 1996]. Le résultat fourni est une liste triée de sous-ensembles disjoints de variables. Cette méthode peut se décomposer en 3 étapes :

- Le calcul et la discrétisation des différents critères pour chaque variable,
- L'application de la méthode d'agrégation d'opinions sur les résultats obtenus à l'étape précédente,
- La recherche du sous-ensemble optimal.

3.1 Calcul et discrétisation des critères

Nous avons sélectionné un ensemble de 10 critères myopes de sélection : l'entropie de Shannon, le gain d'information, le ratio du gain, le gain normalisé, la distance de Mantaras, le critère de Gini, le chi2, le critère de Tschuprow, le coefficient de Cramer et le τ de Zhou.

Les calculs de chaque critère pour la totalité des variables s'effectuent en parallèle. Le résultat obtenu est un ensemble constitué de 10 listes ordonnées dans l'ordre décroissant de l'importance des variables.

Deux variables pouvant être aussi pertinentes l'une que l'autre vis-à-vis de la variable endogène même si elles n'apportent pas le même type d'information, nous introduisons la notion d'équivalence de variables. Afin de définir cette notion, nous considérons un problème d'apprentissage caractérisé par un ensemble d'individus $O = \{o_1, \dots, o_j, \dots, o_n\}$ décrits par un ensemble de variables $X = \{x_1, \dots, x_i, \dots, x_p\}$ nommé l'ensemble initial des variables.

Soit $CR = \{cr_1, \dots, cr_k, \dots, cr_{10}\}$ l'ensemble des 10 critères myopes de sélection, choisis avec $cr_k = \{cr_{k1}, \dots, cr_{ki}, \dots, cr_{kp}\}$, l'ensemble des valeurs du critère k pour les p variables de X.

Les valeurs cr_{ki} de chaque critère sont normalisées à l'aide de la transformation suivante : pour une variable $x_i \in X$ et un critère $cr_k \in CR$, la valeur normalisée du critère est :

$$cr_{ki,N} = \frac{cr_{ki} - \text{Min}(\{cr_k\})}{\text{Max}(\{cr_k\}) - \text{Min}(\{cr_k\})}.$$

Après leur normalisation, ces valeurs sont discrétisées en déciles. La discrétisation permet d'affecter à chaque variable x_i un rang pour chaque critère cr_k de la manière suivante :

- Pour les critères qui doivent être minimisés :
 Si $cr_{ki,N} \in [0; 0.1[$ alors
 $R_{ki} = 1$
 Si $cr_{ki,N} \in [0.1; 0.2[$ alors
 $R_{ki} = 2$
 ...

- Si $cr_{ki,N} \in [0.9; 1]$ alors
 $R_{ki} = 10$
- Pour les critères qui doivent être maximisés :
 Si $cr_{ki,N} \in [0; 0.1[$ alors
 $R_{ki} = 10$
 Si $cr_{ki,N} \in [0.1; 0.2[$ alors
 $R_{ki} = 9$
 \dots
 Si $cr_{ki,N} \in [0.9; 1]$ alors
 $R_{ki} = 1$

R_{ki} est le rang affecté à la variable $x_i \in X$ pour le critère $cr_k \in CR$. La variable la plus pertinente est celle possédant le rang le plus faible.

Ainsi la notion d'équivalence se définit de la manière suivante : deux variables sont équivalentes du point de vue d'un critère particulier si et seulement si pour ce critère, elles ont le même rang : $x_i \Leftrightarrow x_j \Leftrightarrow R_{ki} = R_{kj}$.

3.2 Agrégation des résultats des critères

Pour toutes méthodes d'agrégation [Marcotorchino, 1981], il convient de définir l'ensemble des juges et l'ensemble des individus. Dans notre cas, les individus sont les variables et les juges sont l'ensemble des critères. Nous utilisons la méthode d'agrégation d'opinions développée dans [Nicoloyannis *et al.*, 1999] et [Nicoloyannis *et al.*, 1998] et basée sur [Marcotorchino et Michaud, 1981] et [Michaud, 1982]. Nous n'allons pas décrire en détails cette technique mais en présenter le principe sous-jacent.

Pour tout couple d'individus (x_i, x_j) , chaque juge émet un avis $A_k(i, j)$. A_k , l'opinion du juge k est une application de $X \times X$ dans $\{Pref, NPref, EQ\}$. Ainsi,
 $A_k(i, j) = Pref \Leftrightarrow$ le juge k préfère x_i à $x_j \Leftrightarrow R_{ki} < R_{kj}$
 $A_k(i, j) = NPref \Leftrightarrow$ le juge k préfère x_j à $x_i \Leftrightarrow R_{ki} > R_{kj}$
 $A_k(i, j) = EQ \Leftrightarrow$ le juge k considère x_i et x_j comme équivalentes $\Leftrightarrow R_{ki} = R_{kj}$
 Le résultat que nous désirons obtenir est une opinion OP dite opinion de type préférence large et qui engendre une relation de préordre total sur X . OP est une application de $X \times X$ dans $\{Pref, NPref, EQ\}$.

Définition 1 : Le degré d'accord $\rho_{ij}(OP, A_k)$ entre les avis $OP(i, j)$ et $A_k(i, j)$ est défini comme dans le tableau 1.

Définition 2 : Le degré d'accord entre les opinions OP et A_k est $DA(OP, A_k) = \sum_{(x_i, x_j) \in X} \rho_{ij}(OP, A_k)$.

Définition 3 : Le degré d'accord entre l'opinion OP et l'opinion de tous les juges est $DA(OP) = \sum_{k=1}^{10} DA(OP, A_k)$.

Notre problème consiste donc à construire une opinion OP qui engendre un préordre total sur X et qui maximise $DA(OP)$. Pour la maximisation, la méthode du recuit simulé [Kirkpatrick *et al.*, 1983] est utilisée. Ce problème de programmation linéaire

OP / A_k	$Pref$	$NPref$	EQ
$Pref$	1	0	1/2
$NPref$	0	1	1/2
EQ	1/2	1/2	1

TAB. 1 – Degrés d'accord ρ_{ij} .

NP-difficile peut être résolu par différentes méta-heuristiques ; nous avons sélectionné la méthode du recuit-simulé car elle est facile à mettre en oeuvre et conduit à des temps de calcul de quelques secondes.

Après l'application de cette technique d'agrégation, nous obtenons une liste ordonnée $L = \{l_1, \dots, l_m, \dots, l_M\}$ de sous-ensembles disjoints de variables.

3.3 Découverte du sous-ensemble optimal de variables

Jusqu'à présent, nous étions dans une optique d'approche filtre. Lors de cette étape, nous sommes dans une optique d'approche enveloppe. L'avantage d'utiliser une approche enveloppe est lié au fait que l'influence du sous-ensemble de variables sur les performances de l'algorithme d'apprentissage est prise en compte. Le processus de détermination du sous-ensemble optimal s'effectue de la manière suivante : à la m^{ieme} itération, le sous ensemble de variable $l_m \in L$ est ajouté au sous-ensemble optimal de variables. Le critère d'arrêt est double : il y a arrêt du processus soit lorsque le taux d'erreur est constant sur deux itérations soit lorsque que l'on assiste à une augmentation du taux d'erreur.

Le processus est le suivant :

$X^* = \{\}$

Pour $m = 1$ à M faire

$X^* = X^* \cup l_m$

Application de l'algorithme d'apprentissage

Si taux d'erreur à l'itération $m \geq$ taux d'erreur à l'itération $m - 1$ Alors

Arrêt

$X^* = X^* - l_m$

Fin Pour

Le résultat sera le sous-ensemble optimal X^* .

4 Expérimentations

Les variables quantitatives ont été discrétisées avec la méthode Fusinter. La sélection de variables s'est effectuée sur 30% des individus tout en gardant la répartition initiale des classes. Les tests avec MIFS et ReliefF ont également été effectués sur ces mêmes 30%. Les 70% restant sont utilisés pour la phase d'apprentissage. Pour cela, nous avons choisi une 10-cross validation et les algorithmes d'apprentissage sont : l'arbre d'induction ID3 et les bayésiens naïfs. Les tests avant sélection ont également été effectués sur

ces mêmes 70% de la base. Les bases de test utilisées sont issues de la collection de l'UCI Irvine, (Blake, 98).

Les tableaux 2 et 3 montrent les taux d'erreur et les écarts-type associés obtenus avant et après la sélection de variables respectivement avec ID3 et les bayésiens naïfs. Les résultats obtenus avec ID3 sont intéressants. En effet, excepté pour la base Iono, on assiste à une diminution du taux d'erreur et/ou à une stabilisation des résultats (diminution de l'écart-type). Pour les bayésiens naïfs, les résultats avant et après sélection sont sensiblement identiques. Cependant, le nombre de variables diminue alors que les taux d'erreur ne connaissent pas d'augmentation.

Bases	Avant Sélection		Après Sélection	
	Taux d'erreur	Ecart Type	Taux d'erreur	Ecart Type
Tic Tac Toe	28,44	7,53	26,78	2,65
Breast	5,9	2,64	4,47	2,84
CRX	14,46	5,44	15,7	3,1
Diabetes	24,3	3,97	24,3	4,55
Pima	25,24	5,76	26,73	5,24
Vehicle	30,59	5,54	33,1	5,32
Austra	17,16	6,21	17,56	6,87
Cleve	27,1	9,18	21,9	8,67
Heart	31,05	6,42	26,32	11,04
Iono	10,92	4,37	11,75	4,65
German	29,57	6,13	29,86	4,76

TAB. 2 – *Evaluation de notre méthode de sélection avec ID3.*

Bases	Avant Sélection		Après Sélection	
	Taux d'erreur	Ecart Type	Taux d'erreur	Ecart Type
Tic Tac Toe	28,72	4,1	28,15	6,5
Breast	2,86	1,87	2,65	2,05
CRX	14,65	7,24	15,3	3,75
Diabetes	24,67	2,96	24,48	7,87
Pima	21,52	4,54	22,84	5,95
Vehicle	33,61	5,27	33,95	4,18
Austra	14,65	3,42	15,27	3,61
Cleve	20,04	5,31	17,77	6,14
Heart	16,84	5,16	18,95	7,88
Iono	6,9	5,21	10,58	5,03
German	23,86	2,93	24,43	7,63

TAB. 3 – *Evaluation de notre méthode de sélection avec les Bayésiens Naïfs.*

Les tableaux 4 et 5 indiquent le nombre de variables sélectionnées respectivement avec ID3 et les bayésiens naïfs. A l'exception de la base Tic Tac Toe, le nombre de

Sélection de variables et agrégation d'opinions

variables sélectionnées par notre méthode est inférieur à celui sélectionné par ReliefF et/ou MIFS. Pour les bayésiens naïfs, nos résultats se situent entre ceux de ReliefF et ceux de MIFS.

Bases	Sans Sélection	Notre méthode	ReliefF	MIFS
Tic Tac Toe	9	7	5	3
Breast	9	3	6	9
CRX	15	3	2	7
Diabetes	8	2	4	5
Pima	8	2	7	4
Vehicle	18	14	18	6
Austra	14	1	2	13
Cleve	13	7	6	8
Heart	13	2	2	13
Iono	34	2	25	8
German	20	5	14	3

TAB. 4 – Nombre de variables sélectionnées avec ID3.

Bases	Sans Sélection	Notre méthode	ReliefF	MIFS
Tic Tac Toe	9	7	5	3
Breast	9	7	6	9
CRX	15	5	2	7
Diabetes	8	6	4	5
Pima	8	5	7	4
Vehicle	18	12	18	6
Austra	14	2	2	13
Cleve	13	5	6	8
Heart	13	8	2	13
Iono	34	26	25	8
German	20	9	14	3

TAB. 5 – Nombre de variables sélectionnées avec les Bayésiens Naïfs.

Le tableau 6 montre le nombre d'itérations effectuées lors de la phase enveloppe de notre méthode. Le nombre maximal d'itérations est de l'ordre de 9 (pour la base Vehicle). Le nombre d'appels à l'algorithme d'apprentissage est donc bien plus faible que pour les méthodes enveloppe pures.

Les tables 7 et 8 comparées aux tables 2 et 3 montrent que notre méthode est comparable du point de vue des taux d'erreur à MIFS et ReliefF.

Bases	ID3	BN
Tic Tac Toe	4	4
Breast	3	5
CRX	2	3
Diabetes	3	4
Pima	3	4
Vehicle	9	7
Austra	2	3
Cleve	5	4
Heart	2	4
Iono	3	6
German	5	7

TAB. 6 – *Nombre d'itérations nécessaires à notre méthode.*

5 Conclusion

Dans cet article, nous avons présenté une méthode de sélection de variables pour variables qualitatives, basée sur l'agrégation d'opinion.

C'est une méthode hybride entre approche filtre et enveloppe qui possède les avantages de chaque approche et qui permet de réduire leurs inconvénients :

- L'influence des variables sélectionnées sur l'algorithme d'apprentissage utilisé est pris en compte. Ainsi, les variables sélectionnées sont différentes suivant l'algorithme utilisé.
- Les temps de calcul sont largement inférieurs à ceux des méthodes enveloppe pures grâce à l'utilisation du préordre et à l'obtention d'une liste triée de sous-ensembles de variables.

Du point de vue du nombre de variables sélectionnées, les résultats que nous obtenons sont dans la plupart des cas comparables voire meilleurs à ceux obtenus par ReliefF et MIFS. Du point de vue de la qualité d'apprentissage, nous assistons à une diminution des taux d'erreur après la sélection. Les comparaisons en terme de taux d'erreur entre notre méthode et ReliefF ou MIFS sont probants.

Nous envisageons d'améliorer la méthode proposée suivant divers aspects. La méthode de discrétisation utilisée pour les valeurs des critères doit être plus appropriée. Nous voudrions, également, que le résultat de la méthode d'agrégation ne soit plus une liste de sous-ensembles de variables, mais le sous-ensemble optimal de variables.

Il serait également intéressant de ne plus travailler uniquement sur des variables qualitatives mais sur des variables de tous types (numériques, qualitatives, quantitatives,...).

Sélection de variables et agrégation d'opinions

Bases	ReliefF		MIFS	
	Taux d'erreur	Ecart type	Taux d'erreur	Ecart type
Tic Tac Toe	30,51	5,9	30,81	7,11
Breast	5,29	3,16	5,9	2,64
CRX	17,54	5,88	16,12	6,7
Diabetes	25,78	4,05	23,19	4,9
Pima	25,05	7,69	24,87	4,83
Vehicle	42,25	6,52	40,62	7,39
Austra	15,31	5,23	17,17	4,12
Cleve	40,54	7,77	24,68	10,27
Heart	27,38	9,06	28,42	9,76
Iono	11,78	3,94	15,75	8,71
German	30,14	6,01	27,43	5,06

TAB. 7 – *MIFS et ReliefF avec ID3.*

Bases	ReliefF		MIFS	
	Taux d'erreur	Ecart type	Taux d'erreur	Ecart type
Tic Tac Toe	27,97	4,19	28,87	5,42
Breast	3,45	2,56	2,86	1,87
CRX	16,53	2,8	14,66	5,7
Diabetes	25,81	6,13	21,32	4,43
Pima	25,04	3,41	21,33	4,3
Vehicle	45,82	8,78	39,85	8,01
Austra	15,28	5,15	14,28	3,08
Cleve	40,67	4,33	20,52	11,34
Heart	21,05	10,53	17,89	10,04
Iono	9,32	6,22	5,22	4,4
German	30,71	4,96	26,29	3,63

TAB. 8 – *MIFS et ReliefF avec les Bayésiens naïfs.*

Références

- [Almuallim et G., 1992] H. Almuallim et Dietterich T. G. Efficient algorithms for identifying relevant features. Technical Report 92-30-03, 1992.
- [Battiti, 1994] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. on Neural Networks*, 5:537–550, July 1994.
- [Breiman et al., 1984] L. Breiman, J. H. Friedman, R. A. Olshen, et C. J. Stone. *Classification and Regression trees, The Wadsworth Statistics/Probability Series*, Wadsworth, Belmont, CA. 1984.
- [De Mantaras, 1991] R.L. De Mantaras. A distance-based attribute selection measure for decision tree induction. In *Machine Learning*, volume 6, pages 81–92, 6-9 1991.
- [Hall, 2000] Mark A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. In *Proc. 17th International Conf. on Machine Learning*, pages 359–366. Morgan Kaufmann, San Francisco, CA, 2000.
- [Hart, 1984] A. Hart. Experience in the use of an inductive system in knowledge eng. In M. Bramer, editor, *Research and Development in Expert Systems*. Cambridge Univ. Press, Cambridge, MA,, 1984.
- [John et al., 1994] George H. John, Ron Kohavi, et Karl Pfleger. Irrelevant features and the subset selection problem. In *International Conference on Machine Learning*, pages 121–129, 1994. Journal version in AIJ, available at <http://citeseer.nj.nec.com/13663.html>.
- [Jun et al., 1997] B.H. Jun, C.S. Kim, H.Y. Song, et J. Kim. A new criterion in selection and discretization of attributes for the generation of decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 1997.
- [Kira et Rendell, 1992a] K. Kira et L. A. Rendell. The feature selection problem: Traditional methods and a new algorithm. In MIT Press, editor, *Tenth National Conference on Artificial Intelligence*, pages 129–134, 1992.
- [Kira et Rendell, 1992b] K. Kira et L.A. Rendell. A practical approach to feature selection. In *Proceedings of the Tenth International Conference on Machine Learning*, pages 500–512, 1992.
- [Kirkpatrick et al., 1983] S. Kirkpatrick, C. D. Gelatt, et M. P. Vecchi. Optimisation by simulated annealing. *Science*, 220:671–674, 1983.
- [Lallich et Rakotomalala, 2000] S. Lallich et R. Rakotomalala. Fast feature selection using partial correlation for multi-valued attributes. In *Proc. of the 4th European Conf. on Knowledge Discovery in Databases, PKDD 2000*, pages 221–231, 2000.
- [Liu et al., 1998] Huan Liu, Hiroshi Motoda, et Manoranjan Dash. A monotonic measure for optimal feature selection. In *European Conference on Machine Learning*, pages 101–106, 1998.
- [Liu et Setiono, 1996] Huan Liu et Rudy Setiono. A probabilistic approach to feature selection - a filter solution. In *Int. Conf. on Machine Learning*, pages 319–327, 1996.
- [Marcotorchino et Michaud, 1981] F. Marcotorchino et P. Michaud. Heuristic approach to the similarity aggregation problem. *Methods of Operations Research*, 43:395–404, 1981.

- [Marcotorchino, 1981] J. F. Marcotorchino. Agrégation de similarités en classification automatique. *Thèse de Doctorat d'Etat, Université Paris 6*, 1981.
- [Michaud, 1982] P. Michaud. Agrégation à la majorité 1 : Hommage à Condorcet, étude n°f-051. Technical report, Centre Scientifique IBM-France, 1982.
- [Mingers, 1987] J. Mingers. Expert systems – rule induction with statistical data. In *Journal of the Operational Research Society*, 1987.
- [Nicoloyannis *et al.*, 1998] N. Nicoloyannis, M. Terrenoire, et D. Tounissoux. An optimisation model for aggregating preferences : A simulated annealing approach. *Health and System Science*, 2(1-2):33–44, 1998.
- [Nicoloyannis *et al.*, 1999] N. Nicoloyannis, M. Terrenoire, et D. Tounissoux. Pertinence d'une classification. *Revue Electronique sur l'Apprentissage par les Données*, 3(1):39–49, 1999.
- [Quinlan, 1986] J.R. Quinlan. Introduction of decision trees. In *Machine Learning*, volume 1, pages 81–106, 1986.
- [Shannon, 1948] C.E. Shannon. A mathematical theory of communication. In *Bell System Technical Journal*, 1948.
- [Sheinvald *et al.*, 1990] Sheinvald, Dom, Niblack, et Rendell. A modeling approach to feature selection. In *10th Int. Conf. on Pattern Recognition*, 1990.
- [Zhou et Dillon, 1991] X. Zhou et T.S. Dillon. A statistical-heuristic feature selection criterion for decision tree induction. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 13, pages 834–841, 1991.
- [Zighed *et al.*, 1996] D. A. Zighed, R. Rakotomalala, et S. Rabaséda. A discretization method of continuous attributes in induction graphs. In *13th European Meetings on Cybernetics and System Research*, pages 997–1002, 1996.

Summary

Feature selection enables to reduce the representation space of the data. This process gets more and more important because of databases size increase. Therefore we propose a method for discrete attributes based on preferences aggregation. The result, in form of an ordering of features, is furnished by the aggregation of the results obtained by several short-sighted methods of feature selection.