



**HAL**  
open science

## **GAODE et HAODE: Deux nouveaux classifieurs basés sur AODE afin de traiter variables continues**

M. Julia Flores, José A. Gámez, Ana M. Martínez, José M. Puerta

► **To cite this version:**

M. Julia Flores, José A. Gámez, Ana M. Martínez, José M. Puerta. GAODE et HAODE: Deux nouveaux classifieurs basés sur AODE afin de traiter variables continues. 5èmes Journées Francophones sur les Réseaux Bayésiens (JFRB2010), May 2010, Nantes, France. hal-00466859

**HAL Id: hal-00466859**

**<https://hal.science/hal-00466859>**

Submitted on 19 Apr 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

## GAODE et HAODE:

### Deux nouveaux classifieurs basés sur AODE afin de traiter variables continues

Flores, M. Julia — Gámez, José A. — Martínez, Ana M. — Puerta José M.

*Computing Systems Department, Intelligent Systems and Data Mining group, i3A  
Campus Universitario s/n 02071, Albacete, Espagne  
{julia.flores,jose.gamez,anamaria.martinez,jose.puerta}@uclm.es*

---

*RÉSUMÉ. L'AODE est considéré comme l'un des algorithmes représentatifs les plus intéressants parmi les classifieurs bayésiens, si l'on considère non seulement son taux d'erreur peu élevé mais aussi son efficacité. Jusqu'à présent, pour pouvoir construire un classifieur de type AODE, tous les attributs (ou variables) dans un ensemble de données devaient être nominaux ou bien devaient être des variables préalablement discrétisées. Dans cette étude nous proposons deux approches différentes afin de pouvoir traiter directement des attributs numériques. L'un d'eux utilise des réseaux conditionnels gaussiens pour modéliser un ensemble de données uniquement avec attributs numériques; l'autre (version hybride) discrétise le nœud super-parent dans chaque modèle, et utilise une distribution gaussienne univariée pour estimer les probabilités des attributs numériques, mais des distributions multinomiales pour les attributs à valeurs discrètes. Ces deux classifieurs obtiennent des résultats très compétitifs par rapport à l'AODE.*

*ABSTRACT. AODE (Aggregating One-Dependence Estimators) can be seen as one of the most interesting representatives of the Bayesian classifiers, considering not only the low error rate it provides but also its efficiency. Until now, all the attributes in a dataset have had to be nominal to build an AODE classifier or they have had to be previously discretized. In this paper, we propose two different approaches in order to deal directly with numeric attributes. One of them uses conditional Gaussian networks to model a dataset exclusively with numeric attributes; and the other one (hybrid version) keeps the superparent on each model discrete and uses univariate Gaussians to estimate the probabilities for the numeric attributes and multinomial distributions for the categorical ones. Both of them obtain competitive results compared to AODE.*

*MOTS-CLÉS: Réseaux bayésiens, réseaux conditionnels gaussiens, classification, AODE.*

*KEYWORDS: Bayesian networks, conditional Gaussian networks, classification, AODE.*

---

## 1. Introduction

La classification supervisée est une tâche très courante non seulement dans les applications de *data mining* mais aussi dans nombreux aspects de la vie quotidienne, par exemple la détection de spams dans le courrier électronique, les recommandations pour un produit spécifique basées sur les achats antérieurs, etc. Les classifieurs basés sur les réseaux Bayésiens<sup>1</sup> (RB) (Langley *et al.*, 1992) offrent des avantages significatifs sur d'autres approches pour effectuer les tâches de classification. Ils sont capables de traiter naturellement l'incertitude, et non seulement estimer l'étiquette qui est assignée à chaque objet mais aussi la distribution de probabilité concernant les différentes étiquettes de la variable classe.

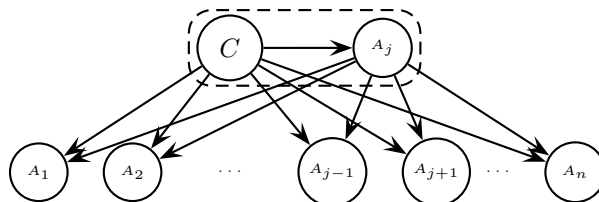
Le classifieur de Bayes naïf (NB du anglais) (Duda *et al.*, 1973) correspond à la structure la plus simple qui soit, en posant l'hypothèse que les attributs  $\{A_1 \dots A_n\}$  sont indépendantes conditionnés à la classe  $C$ . Au cours de ces dernières années, l'attention des chercheurs s'est focalisée sur le développement de variantes de NB pour alléger l'hypothèse d'indépendance entre les attributs et, à ce jour, l'AODE s'est révélée comme l'option la plus attirante à cause de sa capacité à améliorer la précision de NB avec une légère augmentation de la complexité temporelle (de  $\mathcal{O}(n)$  jusqu'à  $\mathcal{O}(n^2)$ ), où  $n$  désigne le nombre d'attributs. Une étude extensive qui compare les différentes techniques Bayes semi-naïves (Zheng *et al.*, 2005) a prouvé que l'AODE est significativement meilleur en termes de réduction de l'erreur quand on le compare avec le reste de techniques semi-naïves.

Jusqu'à présent, la seule façon d'entraîner un classifieur AODE à fonctionner avec un ensemble de données (ED) contenant des attributs numériques à été de les discrétiser avant de construire le modèle, ce qui peut représenter un handicap car ce processus, par définition entraîne une perte intrinsèque d'information. Pourtant, quand on considère les variables numériques, le problème se pose sur la façon de modéliser la distribution de probabilité pour une variable conditionnée, pas seulement par la classe (qui est discrète), mais aussi par une autre attribut numérique. Les réseaux gaussiens (RG) (Geiger *et al.*, 1994) ont été proposés comme étant une bonne alternative à la discrétisation directe des variables continues. Un RG est similaire à un RB, mais cela suppose que tous les attributs sont échantillonnés à partir d'une distribution de densité gaussienne, au lieu d'une distribution multinomiale. Malgré cette forte supposition, les distributions gaussiennes fournissent habituellement des approximations raisonnables pour beaucoup de distributions réelles.

Dans notre étude, deux approches ont été développées pour traiter des variables continues en AODE : GAODE et HAODE (Flores *et al.*, 2009b), les deux héritent de la même structure que l'AODE. Dans la première, nous avons utilisé des réseaux conditionnelles gaussiennes (RCG) pour modéliser la relation avec l'attribut numérique conditionnée à la classe discrète et un autre attribut numérique ; par conséquent, ce classifieur se limite à des EDs numériques. Dans la seconde structure, c'est une version

---

1. Nous l'appellerons *classifieurs bayésienes* pour simplifier.



**Figure 1.** Structure du classifieur AODE.

discrète d'un attribut super parent qui est considéré dans chaque modèle ; donc la relation précédente peut être estimée par une distribution gaussienne univariée. La seconde approche applique des distributions multinomiales pour les variables enfants discrètes, étant capable de traiter directement tous les types de données.

Le présent document est organisé comme suit : les sections 2 et des 3 donnent une présentation d'AODE et RCGs respectivement, lesquels représentent la base principale des nouveaux classifieurs qui ont été développés. La section 4 fournit une explication détaillée des algorithmes conçus. Dans la section 5, nous allons décrire l'organisation des expériences et les résultats. Et finalement, la section 6 résume les conclusions principales de notre travail et présente les perspectives liées à cette étude.

## 2. AODEs : Averaged One-Dependence Estimators

Le classifieur AODE (Webb *et al.*, 2005) est considéré comme une amélioration du classifieur NB et aussi une bonne alternative à d'autres tentatives telles que celles de LBR (Zheng *et al.*, 2000) et SP-TAN (Keogh *et al.*, 1999), car si ces derniers offrent des taux de précision similaires, l'AODE est incontestablement en tête en efficacité au niveau du temps de classification comparé au premier et en temps d'entraînement par rapport au second.

L'AODE fait la moyenne des  $n$  classifieurs SPODE avec chaque attribut différent désigné sous le nom de super-parent (équation 1).

Un SPODE est un estimateur de dépendance-type 1 (ODEs) (Sahami, 1996) où chaque attribut dépend de la classe et d'un autre attribut commun, le super-parent.

$$\operatorname{argmax}_{c \in \Omega_C} \left( \sum_{j=1, N(a_j) > m}^n p(c, a_j) \prod_{i=1, i \neq j}^n p(a_i | c, a_j) \right) \quad [1]$$

où  $C$  représente la classe,  $\Omega_C$  l'ensemble de ses étiquettes,  $A_j$  l'attribut super-parent et  $\{a_1, a_2, \dots, a_n\}$  l'instance à classifier. La condition  $N(a_j) > m$  est utilisée comme un seuil pour éviter prédire à partir d'attributs dotés d'échantillons insuffisants. S'il n'y a pas de valeur qui excède ce seuil, les résultats sont équivalents à NB.

En période d'entraînement, l'AODE a une complexité temporelle  $\mathcal{O}(tn^2)$ , où  $t$  est le nombre d'échantillons d'entraînement ; tandis que la complexité spatiale est  $\mathcal{O}(k(nv)^2)$ ,  $v$  représentant la moyenne du nombre de valeurs par attribut et  $k$  le nombre de classes. La complexité temporelle résultante au temps de classification est  $\mathcal{O}(k(nv)^2)$ , alors que la complexité spatiale est  $\mathcal{O}(k(nv)^2)$ .

### 3. Réseaux Conditionnels Gaussiens (RCGs)

Les nœuds dans un RB peuvent être modélisés par une fonction de distribution gaussienne, que l'on appelle aussi distribution Normale. Toutes les distributions gaussiennes peuvent être définies par deux paramètres : la moyenne ( $\mu$ ) et la variance ( $\sigma^2$ ) respectivement. De même, chaque nœud continu peut avoir une distribution gaussienne pour chaque configuration de ses parents discrets. Si un nœud continu a pour parents un ou plus d'un nœuds continus, la moyenne peut être linéairement dépendante des états de ses parents continus. C'est l'idée fondamentale qui est sous-jacent dans les RCGs (Lauritzen *et al.*, 2001). On doit noter que les nœuds discrets ne peuvent pas avoir de parents continus.

Une propriété remarquable des RCGs est celle d'offrir une ossature où l'exactitude dans l'inférence est garantie. Un autre avantage des RGs est le fait qu'ils ont seulement besoin de  $\mathcal{O}(n^2)$  paramètres pour modéliser un graphe complet.

En général, chaque nœud va enregistrer une fonction de densité locale (modèle de régression linéaire), où la distribution pour une variable continue  $X$  avec des parents discrets  $\mathbf{Y}$  et aussi continus  $\mathbf{Z} = \{Z_1, \dots, Z_s\}$  (étant  $s$  le nombre de parents continus) est une distribution gaussienne unidimensionnelle couvrant les états de ses parents (DeGroot, 1970) :

$$f(X|\mathbf{Y} = y, \mathbf{Z} = z; \Theta) = \mathcal{N}(x : \mu_X(y) + \sum_{j=1}^s b_{XZ_j}(y)(z_j - \mu_{Z_j}(y)), \sigma_{X|\mathbf{Z}}^2(y)) \quad [2]$$

- $\mu_X(y)$  est la moyenne de  $X$  avec la configuration  $Y = y$  de ses parents discrets.
- $\mu_{Z_j}(y)$  est la moyenne de  $Z_j$  avec la configuration  $Y = y$  de ses parents discrets.
- $\sigma_{X|\mathbf{Z}}^2(y)$  est la variance conditionnée à  $X$  sur ses parents continus  $\mathbf{Z}$  et aussi selon la configuration  $Y = y$  de ses parents discrets.
- $b_{XZ_j}(y)$  est un terme de régression qui individuellement mesure la force d'une liaison entre  $X$  et chaque parent continu (ce sera 0 si il n'y a pas un arc entre eux).

Les paramètres locaux sont donnés par  $\Theta = (\mu_X(y), b_X(y), \sigma_{X|\mathbf{Z}}^2(y))$ , où  $b_X(y) = (b_{XZ_1}(y), \dots, b_{XZ_s}(y))^t$  est un vecteur colonne.

Ainsi, si l'on focalise sur le cas bivarié, où la variable  $X$  est uniquement conditionnée par une variable continue  $Z$  plus le variable discrète mentionnée, la variance

conditionnelle et le terme de régression peuvent être facilement obtenus, comme le montrent les équations 3 et 4 <sup>2</sup> :

$$\sigma_{X|Z}^2(y) = \sigma_X^2(y) - b_{XZ}^2(y)\sigma_Z^2(y) \quad [3]$$

$$b_{XZ}(y) = \frac{\sigma_{XZ}(y)}{\sigma_Z^2(y)} \quad [4]$$

#### 4. Nouvelles propositions pour traiter les attributs continus avec AODE

##### 4.1. Le classifieur Gaussian AODE (GAODE)

L'idée sous-jacente de ce classifieur consiste à utiliser RGCs pour traiter les attributs continus dans l'AODE. En fait, puisque la variable classe est discrète, si tous les attributs sont limités à être continus, nous pouvons employer la règle de Bayes pour combiner RBs et RGs afin d'encoder les distributions de probabilité conjointes entre les variables, en se basant sur les indépendances conditionnelles définies par l'AODE.

Dans ce cas particulier d'AODE, la fonction de densité pour chaque attribut prédictif doit être estimée pour un nœud doté d'un seul parent discret, c'est à dire, la classe  $C$  et d'un autre parent continu, qui est l'attribut super-parent dans chaque modèle,  $A_j$ . L'adaptation de la fonction de densité conditionnelle gaussienne (CG) dans l'équation 2 pour ce cas est :

$$f(A_i = a_i | C = c, A_j = a_j) = \mathcal{N}(a_i : \mu_i(c) + b_{ij}(c)(a_j - \mu_j(c)), \sigma_{i|j}^2(c)) \quad [5]$$

La structure bayésienne pour GAODE resterait la même qu'en AODE, et l'on obtient son hypothèse MAP lorsque nous remplaçons les distributions de probabilité multinomiales de l'équation 1 avec la fonction de distribution CG correspondante définie par l'équation 5. Alors que la relation entre chaque attribut prédictif conditionnel à la classe et le super-parent correspondant est modélisé par une distribution CG, la relation entre chaque super-parent et la classe est modélisé par une distribution gaussienne univariée (DGU). Alors, en supposant que toutes les variables prédictives sont continues, GAODE va sélectionner l'étiquette de la classe qui maximise la sommation dans l'équation 6 :

$$\underset{c}{\operatorname{argmax}} \left( \sum_{j=1}^n \mathcal{N}(a_j : \mu_j(c), \sigma_j^2(c)) p(c) \cdot \prod_{i=1 \wedge i \neq j}^n \mathcal{N}(a_i : \mu_i(c) + b_{ij}(c)(a_j - \mu_j(c)), \sigma_{i|j}^2(c)) \right) \quad [6]$$

2. Les estimations dans les équations 3 et 4 ont été obtenues en travaillant avec la valeur de  $\sigma_{X|Z}^2(y)$  et  $b_{XZ}(y)$  quand l'inverse de la matrice de précision ( $W^{-1}$ ) et la matrice de covariance ( $\Sigma$ ) à partir du RG sont appariées :

$$W^{-1} = \begin{pmatrix} \sigma_Z^2(y) & b_{XZ}(y)\sigma_Z^2(y) \\ b_{XZ}(y)\sigma_Z^2(y) & \sigma_{X|Z}^2(y) + b_{XZ}^2(y)\sigma_Z^2(y) \end{pmatrix} = \begin{pmatrix} \sigma_Z^2(y) & \sigma_{XZ}(y) \\ \sigma_{XZ}(y) & \sigma_X^2(y) \end{pmatrix} = \Sigma$$

Comme nous pouvons le déduire à partir de notre définition du classifieur qui utilise CGNs, il n'est pas possible de définir la fonction de probabilité correspondante pour une variable discrète conditionnée à un attribut numérique. Tout comme pour l'AODE, tous les attributs jouent un rôle de super-parent dans un modèle, aucun des attributs enfants ne peuvent être discrets et par conséquent, GAODE n'est défini que pour traiter les bases de données qui sont formées exclusivement d'attributs numériques.

Dans ce cas, la complexité spatiale pour *entraîner* et *classifier* va devenir indépendante du nombre des valeur par attribut  $v$ , et c'est  $\mathcal{O}(kn^2)$ . D'ailleurs, puisque le nombre des paramètres nécessaires est indépendant de  $v$ , l'estimation des probabilités est plus fiable par rapport à la version multinomiale dû au fait qu'elles sont modélisées à partir d'un plus grand nombre d'échantillons, surtout quand la taille des tableaux de probabilité conditionnés (CPTs en anglais) est très grande.

La complexité temporelle ne subit aucune variation car les paramètres des distributions gaussiennes et CGs peuvent être calculés de façon incrémentielle.

#### 4.2. Le classifieur Hybrid AODE (HAODE)

Comme nous l'avons vu, le classifieur GAODE défini ne peut traiter que des bases de données contenant uniquement des attributs continus. Afin d'inclure la possibilité d'utiliser toutes sortes de bases de données, nous avons décidé de considérer chaque super-parent comme discret dans son modèle correspondant, en principe, par le biais de n'importe quelle méthode de discretisation. Cependant, seul le super-parent sera discrétisé, pour le reste des attributs, on considèrera leur valeur numérique. De cette façon, ce n'est pas nécessaire d'avoir recours aux RCGs, puisque tous les parents dans le réseau sont discrets, mais en même temps nous conservons la majeure partie de la précision originelle des données numériques.

On peut voir cette procédure comme une manière encore plus simple de résoudre le problème du traitement des super-parents continus pour chaque modèle de l'AODE, puisque nous n'avons pas besoin d'employer RCGs, mais seulement des DUGs.

D'où, l'équation 1 qui est développée de la façon suivante :

$$\operatorname{argmax}_c \left( \sum_{j=1, N(a_j) > m}^n p(a_j, c) \prod_{i=1 \wedge i \neq j}^n \mathcal{N}(a_i : \mu_i(c, a_j), \sigma_i^2(c, a_j)) \right) \quad [7]$$

Cela veut dire que la relation entre le super-parent et la classe est modélisée avec une distribution de probabilité multinomiale, alors que les autres relations, où chaque attribut est conditionné à la classe et au super-parent, sont modélisés par des distributions gaussiennes normales, pourvu qu'ils soient continus.

Comme on l'a déjà noté, ce nouveau classifieur offre un avantage supplémentaire : la capacité de traiter avec tous les types de données, les hybrides (mélanges de variables discrètes et continues) aussi. Dans les cas particuliers où l'attribut enfant est

discret, on utilisera comme AODE une distribution multinomiale. Cette caractéristique représente un atout important en ce qui concerne l'utilisation des RGCs proposés par GAODE, ainsi qu'une simplification évidente dans le calcul des paramètres.

Les modèles construits sont encore 1-dépendant, et c'est la raison pour laquelle les CPTs qui gardent les distributions de probabilité, lorsque cela est nécessaire pour l'HAODE, restent tridimensionnelles, comme pour l'AODE. Dans ce cas, la complexité spatiale va augmenter avec le nombre variables discrètes dans la base de données, le limite supérieure étant le même que pour l'AODE ( $\mathcal{O}(k(nv)^2)$ ).

## 5. Méthodologie expérimentale et résultats

### 5.1. Ensemble de données contenant exclusivement des attributs numériques

Afin d'évaluer la performance des deux classifieurs développés, nous avons effectué des expériences sur un total de 26 EDs numériques, précisément tous les EDs présents sur la page d'accueil de l'Université de Waikato<sup>3</sup>, ayant pour origine le répertoire UCI (Asuncion *et al.*, 2007), qui ciblent les problèmes de classification et ne contiennent que des attributs numériques selon Weka (Witten *et al.*, 2005). Le tableau 1 affiche ces EDs et leurs principales caractéristiques.

**Tableau 1.** Principales caractéristiques des EDs : nombre de variables prédictives ( $n$ ), nombre de classes ( $k$ ) et nombre de cas ( $i$ ).

Id	Datasets	$n$	$k$	$I$	Id	Datasets	$n$	$k$	$I$
1	balance-scale	4	3	625	14	mfeat-fourier	76	10	2000
2	breast-w	9	2	699	15	mfeat-karh	64	10	2000
3	diabetes	8	2	768	16	mfeat-morph	6	10	2000
4	ecoli	7	8	336	17	mfeat-zernike	47	10	2000
5	glass	9	7	214	18	optdigits	64	9	5620
6	hayes-roth	4	4	160	19	page-blocks	10	5	5473
7	heart-statlog	13	2	270	20	pendigits	16	9	10992
8	ionosphere	34	2	351	21	segment	19	7	2310
9	iris	4	3	150	22	sonar	60	2	208
10	kdd-JapanV	14	9	9961	23	spambase	57	2	4601
11	letter	16	26	20000	24	vehicle	18	4	946
12	liver-disorders	6	2	345	25	waveform-5000	40	3	5000
13	mfeat-factors	216	10	2000	26	wine	13	3	178

Le tableau 2 montre les résultats de précision obtenus en utilisant une validation croisée (CV) 5x2 afin d'évaluer les différents classifieurs. Chaque valeur représente la moyenne arithmétique des 10 exécutions. La balle à côté de certaines sorties signifie que le classifieur correspondant à cet ensemble particulier de données, soit obtient le degré de précision le plus élevé ou bien n'est pas significativement pire que le classifieur qui l'obtient. Les résultats ont été comparés en utilisant le F test 5x2 cv défini par (Alpaydin, 1999), qui comporte un taux d'erreur de type I plus bas et une puissance plus élevée que le t-test 5x2 cv. Le niveau de signification a été fixé à 95% ( $\alpha = 0.05$ ).

En plus du GAODE et de l'HAODE, 3 autres classifieurs ont été inclus dans la comparaison. De gauche à droite : classifieur NB à distributions gaussiennes

3. <http://www.cs.waikato.ac.nz/ml/weka/>



**Tableau 2.** Résultats de la précision obtenus pour le NB à distributions gaussiennes (NB-G), NB, AODE ( $m = 1$ ), GAODE et HAODE ( $m = 1$ ) sur les EDs numériques.

Id	NB-G	NB	AODE	GAODE	HAODE
1	●88, 864	77, 632	76, 992	●89, 088	●87, 68
2	●96, 0801	●97, 1102	●96, 6237	●95, 9662	●95, 0787
3	●74, 974	●74, 6875	●74, 5573	●74, 7917	●75, 9115
4	●83, 9881	●80, 7738	●81, 0119	●84, 5238	●84, 3452
5	49, 7196	●60	●60, 7477	●52, 8037	●60, 6542
6	●65, 375	●57, 5	●57, 5	●65, 625	●68, 5
7	●83, 4815	●81, 2593	●80, 8148	●83, 7778	●83, 037
8	●82, 963	●88, 8889	●90, 7123	●92, 0228	●91, 7379
9	●95, 0667	●93, 4667	●93, 3333	●97, 4667	●95, 6
10	85, 7444	84, 5758	90, 3885	91, 8442	●93, 9966
11	64, 06	73, 296	●86, 292	71, 235	●86, 138
12	●54, 2609	●58, 6087	●58, 6087	●57, 3333	●54, 2029
13	92, 29	92, 36	●96, 08	●95, 94	●96, 31
14	75, 7	75, 87	79, 25	●79, 39	●80, 69
15	93, 16	90, 48	●93, 83	●96, 15	●95, 92
16	●69, 32	68, 03	68, 9	●70, 79	●69, 95
17	72, 99	70, 21	74, 63	●77, 42	●78, 1
18	91, 1317	91, 7544	●96, 3167	93, 637	●96, 9181
19	●87, 7142	93, 1336	●96, 6307	●90, 9446	●91, 8144
20	85, 7041	87, 3362	●97, 1161	94, 2085	●97, 5182
21	80, 6753	90, 4416	●94, 1732	86, 6667	●95, 1602
22	67, 5	●75, 6731	●75, 5769	●71, 4423	●75, 9615
23	79, 5131	89, 8544	●92, 7277	79, 8566	77, 3658
24	43, 1678	58, 6052	67, 4704	●68, 5106	●72, 9787
25	80	79, 968	●84, 508	●4, 46	●84, 22
26	97, 4157	96, 9663	●96, 9663	●98, 427	97, 4157
<i>Av</i>	78, 4842	80, 3262	83, 1445	82, 4739	84, 1233

pour traiter les attributs continus (NB-G); et classifieurs NB et AODE avec les EDs précédemment discretisés (NB et AODE) en utilisant la méthode de Fayyad et Irani (MDL) (Fayyad *et al.*, 1993). La discrétisation des attributs super-parents pour HAODE, a également été effectuée à l'aide de cette méthode <sup>4</sup>.

Le tableau 3 montre, dans la moitié supérieure de chaque cellule, la comparaison entre chaque paire d'algorithmes, où chaque entrée  $w-t-l$  à la ligne  $i$  et la colonne  $j$  signifie que l'algorithme à la ligne  $i$  gagne en  $w$  EDs, égalise en  $t$  (c'est à dire, qu'il n'y a pas de différence statistique selon le F test 5x2 cv) et il perd en  $l$  EDs, par rapport à l'algorithme dans la colonne  $j$ . La moitié inférieure de chaque cellule contient les résultats des tests de Wilcoxon, avec  $\alpha = 0,05$ , qui comparent chaque paire d'algorithmes avec les 26 EDs : chaque fois que le résultat du test représente une amélioration significative en faveur d'un classifieur sur l'autre, le nom du gagnant est affiché, sinon N0 est affichée.

En termes de moyenne arithmétique obtenue, on pourrait penser que NB avec discrétisation fonctionne mieux que NB-G, mais le nombre d'EDs où NB-G n'est pas significativement plus mauvais que la meilleure méthode ou bien en fait représente la meilleure méthode, est de 11, contre 10 pour NB. En fait, le test de Wilcoxon ne fait apparaître aucune différence significative entre ces deux méthodes pour ces

4. D'autres expériences ont été réalisées avec les méthodes de discrétisation différentes, et les résultats obtenus suivent la même tendance (Flores *et al.*, 2009a).

**Tableau 3.** Comparaison de la précision entre les paires d'algorithmes.

Ftest Wilcoxon	NB-G	NB	AODE	GAODE
NB	7-16-3			
	NO			
AODE	11-14-1	14-12-0		
	AODE	AODE		
GAODE	12-14-0	12-12-2	5-16-5	
	GAODE	GAODE	NO	
HAODE	13-13-0	13-12-1	6-19-1	6-18-2
	HAODE	HAODE	HAODE	HAODE

EDs. On pourrait espérer que le même raisonnement s'applique à la comparaison entre AODE et GAODE. Toutefois, cela n'est pas tout à fait vrai car la différence entre les moyennes des algorithmes est plus faible et le nombre d'EDs où les résultats ne sont pas sensiblement plus mauvais que pour les autres classifieurs est exactement le même. Dans ce cas, le test de Wilcoxon n'a pas non plus réussi à montrer une différence significative.

Si on analyse ces résultats, nous pouvons confirmer que GAODE et HAODE à la fois, sont significativement meilleurs que NB dans toutes ses versions. En ce qui concerne HAODE, non seulement il obtient la plus grande moyenne de précision, mais aussi le plus grand nombre d'EDs dont les précisions ne sont pas significativement différentes des meilleures fournies par n'importe lequel des autres classifieurs. De plus, selon le test de Wilcoxon, il est significativement meilleur qu'AODE et GAODE en ce qui concerne ce groupe d'EDs, malgré le nombre considérable d'égalisations.

De plus, un test de Friedman a été effectué sur les 5 classifieurs, révélant une différence statistique. Les tests Nemenyi postérieurs (Demšar, 2006; García *et al.*, 2009) ont uniquement rejeté l'hypothèse que les 2 algorithmes ne sont pas significativement différents en faveur de GAODE et HAODE par rapport à NB-G et NB, alors qu'AODE n'a pas pu être révélée significativement mieux que n'importe quel d'entre eux.

## 5.2. Ensembles de données hybrides

Jusqu'à présent, nous avons vu la grande capacité de l'HAODE pour servir d'alternative à l'AODE dans les EDs numériques. Par opposition à GAODE, HAODE est capable de traiter tous les types d'EDs, nous avons aussi effectué des expériences sur 16 EDs hybrides inclus dans un groupe standard de 36 EDs du répertoire UCI, dont les principales caractéristiques sont résumées dans le tableau 4. Tous les EDs numériques dans ce groupe ont été inclus dans la précédente série d'expériences et pour les EDs discrètes les classifieurs sont égaux. C'est la raison pour laquelle dans ce bloc, nous nous concentrons uniquement sur les hybrides.

Le tableau 5 montre les résultats de la précision pour NB (estimation gaussienne ou multinomiaux selon le type des attributs), AODE et HAODE avec un 5x2cv pour l'évaluation et l'application de la méthode de discrétisation déjà mentionnée. La raison pour laquelle l'ordre des EDs a été modifié sera donnée ci-dessous.

**Tableau 4.** *Caractéristiques des EDs hybrides : nombre d'attributs ( $n$ ), nombre de classes ( $k$ ), nombre de cas ( $i$ ), nombre d'attributs discrets et continus ( $\#D$  et  $\#C$ ) et % des valeurs manquantes ( $\%M$ ).*

Id.	Dataset	$n$	$k$	$I$	$\#D$	$\#C$	$\%M$
1	anneal.ORIG	38	6	898	29	9	63,32
2	anneal	38	6	898	29	9	0
3	autos	25	7	205	10	15	11,06
4	colic.ORIG	27	2	368	20	7	18,7
5	colic	27	2	368	15	7	22,77
6	credit-a	15	2	690	9	6	5
7	credit-g	20	2	1000	13	7	0
8	heart-c	13	2	303	7	6	0,17
9	heart-h	13	2	294	7	6	19
10	hepatitis	19	2	155	13	6	5,39
11	hypothyroid	29	4	3772	22	7	5,4
12	labor	16	2	57	8	8	0
13	lymph	18	4	148	15	3	0
14	sick	29	2	3772	22	7	5,4
15	vowel	13	11	990	10	3	0
16	zoo	17	7	101	16	1	0

**Tableau 5.** *Résultats de la précision obtenue avec le NB, AODE et HAODE dans les EDs hybrides.*

Id	NB	AODE	HAODE	$\%M$
16	●90,495	●91,6832	●94,2574	0
13	●81,0811	●80,8108	●82,5676	0
15	50,6667	61,0505	●78,4444	0
7	●74,16	●74,44	●75,32	0
12	●88,4211	●87,7193	●88,0702	0
2	●95,1448	●96,7483	●92,784	0
8	●83,3003	●83,3003	●83,7624	0,17
6	●86,029	●86,2609	78,8696	5
10	●82,3226	●83,0968	●84,3871	5,39
11	●97,7253	●98,0011	95,6416	5,4
14	97,0891	●97,2057	94,5652	5,4
3	●58,7317	●64,1951	●57,561	11,06
4	●69,6196	●69,7826	60,8696	18,7
9	●83,8776	●83,9456	●83,4014	19
5	●79,3478	●81,087	●78,8043	22,77
1	●93,1403	●93,9866	88,7751	63,32
<i>Av</i>	●81,947	●83,3321	●82,3801	

Si chaque  $w-t-l$  notation signifie que HAODE gagne en  $w$  EDs, égalise en  $t$  et perd en  $l$ , si l'on compare avec AODE à 95% niveau de confiance, le classifieur hybride améliore de manière significative sur AODE dans 1 ED, perd dans 5 autres et égalise dans 10 (1-10-5). Même si ce ne sont pas les résultats escomptés, si l'on considère seulement ces EDs hybrides, on ne peut pas prouver qu'il existe un avantage significatif du AODE sur HAODE, car Wilcoxon ne garantit pas de différence statistique.

Recherchant une raison plausible pour expliquer ce fait, et considérant les bons résultats obtenus par HAODE vs AODE dans des EDs numériques (tableau 2), nous avons analysé le pourcentage de variables numériques par rapport aux variables discrètes dans les EDs hybrides, mais aucun modèle significatif n'a été trouvé. Puis, nous nous sommes dirigés vers l'étude de l'impact des valeurs manquantes et, dans ce cas, on peut obtenir une tendance pertinente : *la présence de valeurs manquantes pénalise*

*l'HAODE par rapport à l'AODE.* Ainsi, dans le tableau 5, les EDs hybrides ont été classés selon leur pourcentage de valeurs manquantes. Au dessus de la ligne dans le tableau 5, nous avons ceux qui n'ont presque pas de valeurs manquantes. En fait, le test de Wilcoxon révèle des différences statistiques lorsqu'on considère uniquement les EDs avec des valeurs manquantes. Sur la base de la tendance apparente de l'HAODE à pénaliser les EDs avec des valeurs manquantes, nous avons prétraité tous les EDs avec un filtre non-supervisé pour remplacer les valeurs manquantes avec les modes et moyennes des données existantes dans la colonne correspondante. Les mêmes expériences ont été faites avec des résultats de 2-12-2. Pour le premier groupe d'EDs numériques, les résultats sont les mêmes, parce que seulement le breast-w a 0, 23% de valeurs manquantes. Ces résultats nous amènent à la conclusion que l'HAODE peut être plus sensible aux valeurs manquantes que les autres classifieurs inclus dans la comparaison. Il semble que l'emploi répété de différents estimateurs (moyenne de  $n$  modèles) à partir de peu de données sur les RGs est plus préjudiciable que lorsqu'ils sont faits à partir de multinomiaux.

## 6. Conclusions et perspectives

Dans cet article, nous avons proposé deux variantes pour le classifieur AODE afin de traiter les attributs continus sans effectuer un processus de discrétisation direct sur les EDs. Le premier classifieur, GAODE, applique les RCGs pour modéliser les relations entre chaque attribut prédictif avec ses parents, et il obtient des résultats compétitifs par rapport à l'AODE. Le GAODE implique une réduction de la complexité de l'espace et les paramètres peuvent être informatisés *a priori* en un seul passage sur les données, en conservant ainsi la complexité temporelle de l'AODE. Cette approche peut également fournir un calcul plus fiable et robuste des statistiques nécessaires puisque les paramètres sont exclusivement conditionnés à la classe.

En outre, nous avons présenté un classifieur "hybride", HAODE, qui maintient discret l'attribut super-parent dans chaque modèle. Cette approche offre l'avantage évident de traiter avec toutes sortes d'ensembles de données. Néanmoins, même s'il est en général compétitif par rapport à l'AODE, il a montré une préférence nette pour les EDs de données dotés d'attributs continus et sans valeurs manquantes, où il est significativement meilleure que l'AODE. Le traitement adéquat de ces valeurs manquantes en HAODE dépasse la portée de ce document, donc nous allons l'aborder dans de prochains travaux.

Même si les RGs offrent souvent une approximation raisonnable du monde réel, ils assument que les variables sont échantillonnées à partir de distributions gaussiennes. À l'avenir, nous envisageons d'explorer des distributions plus générales, en particulier l'application des Mélanges d'Exponentielles Tronqués pour AODE, qui impliquent une estimation plus précise, et sont capables de modéliser les réseaux bayésiennes, gaussiens et hybrides.

## Remerciements

Ce travail a été soutenu par la *Consejería de Educación y Ciencia (JCCM)* sous le Projet PCI08-0048-8577574, le Ministère Espagnol d'Education sous le Projet TIN2007-67418-C03-01 et la Bourse FPU avec numéro de référence AP2007-02736.

## 7. Bibliographie

- Alpaydin E., « Combined 5 x 2 cv F test for comparing supervised classification learning algorithms. », *Neural Comput.*, vol. 11, n° 8, p. 1885-1892, November, 1999.
- Asuncion A., Newman D., « UCI Machine Learning Repository », 2007. University of California, Irvine, School of Information and Computer Sciences. <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- DeGroot M. H., *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.
- Demšar J., « Statistical Comparisons of Classifiers over Multiple Data Sets », *J. Mach. Learn. Res.*, vol. 7, p. 1-30, 2006.
- Duda R. O., Hart P. E., Stork D. G., *Pattern classification and scene analysis*, Wiley NY, 1973.
- Fayyad U. M., Irani K. B., « Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning », *Proc. of the 13th Int. Joint Conf. on Artificial Intelligence*, p. 1022-1027, 1993.
- Flores M. J., Gámez J. A., Martínez A. M., Puerta J. M., « Estudio y Comparativa de Diferentes Discretizaciones en Clasificadores Bayesianos », *CAEPIA'09 : Thirteenth Spanish Conference for Artificial Intelligence, Sevilla, Spain, November 9-14*, p. 265-274, 2009a.
- Flores M. J., Gámez J. A., Martínez A. M., Puerta J. M., « GAODE and HAODE : two proposals based on AODE to deal with continuous variables », in A. P. Danyluk, L. Bottou, M. L. Littman (eds), *ICML*, vol. 382 of *ACM International Conference Proceeding Series*, ACM, p. 40, 2009b.
- García S., Herrera F., « An Extension on “Statistical Comparisons of Classifiers over Multiple Data Sets” for all Pairwise Comparisons », *J. Mach. Learn. Res.*, vol. 9, p. 2677-2694, 2009.
- Geiger D., Heckerman D., « Learning Gaussian Networks », *Proc. of the 10th Annual Conf. on Uncertainty in AI*, p. 235-243, 1994.
- Keogh E., Pazzani M., « Learning Augmented Bayesian Classifiers : A Comparison of Distribution-based and Classification-based Approaches », *Proc. of the 7th Int. Workshop on AI and Statistics*, p. 225-230, 1999.
- Langley P., Iba W., Thompson K., « An Analysis of Bayesian Classifiers », *Proc. of the 10th National Conf. on AI*, p. 223-228, 1992.
- Lauritzen S. L., Jensen F., « Stable local computation with conditional Gaussian distributions », *Statistics and Computing*, vol. 11, n° 2, p. 191-203, 2001.
- Sahami M., « Learning limited dependence Bayesian classifiers », *Proc. of the 2nd Int. Conf. on Knowledge Discovery in Databases*, p. 335-338, 1996.
- Webb G. I., Boughton J. R., Wang Z., « Not So Naive Bayes : Aggregating One-Dependence Estimators », *Mach. Learn.*, vol. 58, n° 1, p. 5-24, 2005.
- Witten I. H., Frank E., *Data Mining : Practical Machine Learning Tools and Techniques*, 2 edn, Morgan Kaufmann, 2005.
- Zheng F., Webb G., « A Comparative Study of Semi-naive Bayes Methods in Classification Learning », in S. Simoff, G. Williams, J. Galloway, I. Kolyshkina (eds), *Proc. of the 4th Australian Data Mining Conf.*, p. 141-156, 2005.
- Zheng Z., Webb G. I., « Lazy Learning of Bayesian Rules », *Mach. Learn.*, vol. 41, n° 1, p. 53-84, 2000.