



**HAL**  
open science

## Apport de la négation pour la classification supervisée à l'aide d'associations

François Rioult, Bruno Zanuttini, Bruno Crémilleux

► **To cite this version:**

François Rioult, Bruno Zanuttini, Bruno Crémilleux. Apport de la négation pour la classification supervisée à l'aide d'associations. Conférence francophone sur l'apprentissage automatique (CAp 2008), 2008, France. pp.183-196. hal-00995242

**HAL Id: hal-00995242**

**<https://hal.science/hal-00995242>**

Submitted on 23 May 2014

**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.

# Apport de la négation pour la classification supervisée à l'aide d'associations

François Rioult, Bruno Zanuttini, Bruno Crémilleux

GREYC, CNRS - UMR 6072, Université de Caen  
F-14032 Caen cedex {Prenom.Nom}@info.unicaen.fr

**Résumé** : L'utilisation de règles d'association est une méthode bien connue pour des tâches de classification supervisée. Traditionnellement, les règles utilisées sont de la forme  $X \rightarrow c$ , où  $X$  est un ensemble d'attributs et  $c$  est une valeur de classe. Nous nous intéressons à la généralisation naturelle de ces règles consistant à autoriser des attributs négatifs en prémisse ainsi que des valeurs de classe négatives en conclusion. Comprendre l'impact des règles avec négation dans un processus de classification est une tâche cruciale. Nous proposons un classifieur utilisant de telles règles et étudions l'apport de ces dernières pour la tâche de classification. Contrairement à l'intuition, il s'avère que, toutes choses égales par ailleurs, l'utilisation de ces règles avec négation en classification est délicate.

**Mots-clés** : Classification supervisée, règles d'association, règles généralisées avec négation.

## 1 Introduction

Les méthodes de classification à base de règles d'association sont nombreuses, populaires, et ont donné lieu à de nombreux travaux originaux pour optimiser le processus. La qualité des résultats (score de classification de 85% en moyenne sur une vingtaine de benchmarks de l'UCI) est telle qu'il devient difficile d'améliorer les résultats : les propositions récentes sont de plus en plus complexes et leur amélioration semble difficile.

À peu d'exceptions près [Antonie & Zaïane (2004); Baralis & Garza (2006)], ces méthodes utilisent des règles d'association *classiques* (ou *positives*), c'est-à-dire de la forme  $X \rightarrow c$ , où  $X$  est un ensemble d'attributs et  $c$ , une valeur de classe. Pour un objet à classer contenant les attributs de l'ensemble  $X$ , cette règle prescrit la classe  $c$ . Les classifieurs utilisant ces règles sont construits en extrayant les règles ayant un certain support et une certaine confiance sur un jeu d'apprentissage, puis en les faisant voter pour un nouvel objet à classer.

Une généralisation naturelle de ces règles consiste à autoriser des *valeurs de classe négatives* en conclusion. Une règle peut alors être de la forme  $X \rightarrow \bar{c}$ , excluant la classe  $c$  pour un objet contenant les attributs de  $X$ . Une autre généralisation naturelle consiste à autoriser des *attributs négatifs* en prémisse. Une règle peut alors être de la

forme  $X\bar{Y} \rightarrow c$ , prescrivant la classe  $c$  pour un objet contenant tous les attributs de  $X$  et aucun attribut de  $Y$ . Bien entendu, ces deux généralisations peuvent également être combinées. D'un point de vue logique, on peut voir les règles classiques comme des règles de Horn, et les règles d'association généralisées comme des règles quelconques.

L'intuition suggère que les règles généralisées proposent une sémantique étendue par rapport aux règles traditionnelles, et que les classifieurs puissent en bénéficier. Nous montrons donc comment construire un classifieur utilisant des règles généralisées. Nous adaptons pour cela la méthode de référence CMAR [Li *et al.* (2001)] en ne modifiant que la nature des règles utilisées, afin d'évaluer l'impact des règles généralisées sur la classification.

Comme dans les approches à base de règles classiques, la construction du classifieur passe par l'extraction des règles non redondantes selon une contrainte de fréquence. Nous commençons par montrer comment obtenir des règles généralisées à prémisse et conclusion minimales pour un même support puis nous expliquons comment adapter CMAR pour exploiter ces règles.

Ces généralisations systématiques des techniques utilisées avec les règles classiques nous permettent de mesurer l'apport de la négation pour la classification supervisée, toutes choses égales par ailleurs. Nous utilisons à cette fin les jeux de référence de l'UCI [Blake & Merz (1998)].

Le plan de l'article est le suivant. Nous rappelons tout d'abord la définition des règles d'association, revoyons les méthodes de classification supervisée qui utilisent ces règles et introduisons la généralisation des règles d'association (section 2). Nous présentons ensuite une méthode d'extraction de règles non redondantes et sa déclinaison en méthode de classification supervisée ainsi que les scores obtenus (section 3). Ce classifieur permet de mesurer expérimentalement l'apport de la négation (section 4) pour la classification à base d'associations et nous concluons à la section 5.

## 2 Règles d'association et classification supervisée

Les préliminaires qui suivent permettent de situer le contexte de cet article. Ils présentent les règles d'association et les méthodes de classification supervisée qui les utilisent. Nous terminons par la définition des règles d'association généralisées.

### 2.1 Bases et motifs

Nous nous plaçons dans le cadre classique de bases de données *transactionnelles*. Une telle base est donnée par un *contexte booléen*  $r = (\mathcal{A}, \mathcal{O}, R)$ , où  $\mathcal{O}$  est l'ensemble des objets étudiés,  $\mathcal{A}$  est l'ensemble des attributs de ces objets, et  $R$  est une relation binaire qui indique la présence ou l'absence de chaque attribut binaire dans chaque objet. La relation  $R$  peut être représentée par une matrice booléenne, mais nous considérerons également chaque objet comme un sous-ensemble de  $\mathcal{A}$  (par exemple,  $o_1 = \{a_1, a_3, a_5\}$ , que l'on notera sous forme de chaîne  $a_1a_3a_5$ ). La table 1 donne un exemple d'une telle base.

Un *motif*  $X$  est un sous ensemble de  $\mathcal{A}$ , son *support* dans  $r$  est l'ensemble des objets qui le contiennent (nous noterons  $supp(X) = \{o \in \mathcal{O} \mid X \subseteq o\}$ ) et sa *fréquence*

objets	attributs						
	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$
$o_1$	×		×		×		
$o_2$		×	×		×		
$o_3$	×		×		×		
$o_4$	×			×		×	
$o_5$		×	×			×	
$o_6$		×	×			×	
$o_7$	×			×			×
$o_8$		×		×			×

TAB. 1 – Exemple d’un contexte booléen  $r$ .

$\mathcal{F}(X) = |supp(X)|$  est le nombre d’objets du support. Par exemple,  $supp(a_1a_3a_5) = o_1o_3$  et sa fréquence vaut 2 dans la table 1.

## 2.2 Règle d’association

Une règle d’association *classique* [Agrawal & Srikant (1994)] est une expression  $X \rightarrow Y$ , où  $X$  et  $Y$  sont deux motifs disjoints. Une telle règle est quantifiée par une *fréquence*, définie comme celle de  $X \cup Y$ , et une *confiance*  $conf(X \rightarrow Y) = \mathcal{F}(X \cup Y) / \mathcal{F}(X)$ . La fréquence représente le nombre d’objets auxquels la règle s’applique, et la confiance représente la probabilité conditionnelle de présence de  $Y$  dans les objets contenant  $X$ . Sur notre exemple, la règle  $a_1a_3 \rightarrow a_5$  a une fréquence de 2 et une confiance de 1,  $a_2a_3 \rightarrow a_6$  a une fréquence de 2 et une confiance de  $2/3$ .

## 2.3 Classification à base de règles d’association

Cette section rappelle les grands principes de la classification supervisée à base d’associations. On suppose travailler sur des données *supervisées*, réparties en classes à l’aide d’*étiquettes*. Un attribut  $C$  spécifique, de valeurs  $c_1, \dots, c_n$ , indique pour chaque objet (ou *instance*) à quelle *classe* il appartient (un objet appartient exactement à une classe). Sur l’exemple précédent, on pourra considérer que l’attribut  $a_1$  désigne la première classe, et  $a_2$  la seconde.

Étant donnée une telle base (d’apprentissage), la *classification supervisée* consiste à construire un *classifieur*, qui attribuera correctement une classe à un objet nouveau.

Typiquement, plusieurs règles peuvent être déclenchées pour un même objet à classer, potentiellement avec des conclusions divergentes. On utilise alors un schéma de vote entre ces règles, en pondérant l’importance de chacune par une mesure comme le  $\chi^2$  entre la prémisse et la conclusion [Li *et al.* (2001)].

De façon générale, trois problèmes se posent lors de la mise au point d’un classifieur. Premièrement, il faut éviter le sur-apprentissage, qui spécialise excessivement le modèle sur l’ensemble d’apprentissage. Dans ce cas, le classifieur reconnaît parfaitement les objets qu’il a appris, mais n’est pas efficace sur les objets nouveaux. Dans le cas des

règles d'association, on se limite aux règles dont la fréquence dépasse un seuil fixé afin d'éviter une surabondance de règles trop spécialisées. Le sur-apprentissage est aussi limité par l'emploi de règles à prémisses minimales.

Le second problème est que les classes ne sont pas toujours de populations homogènes. Le classifieur appris risque alors de se spécialiser sur la classe prédominante. Dans le cas des règles d'association, un remède consiste à définir une notion de couverture des règles relativement à l'ensemble d'apprentissage. Les règles sont choisies selon une mesure et sélectionnées lorsqu'elles classent un objet à couvrir. Lorsqu'un objet est classé par exemple par au moins trois règles, il n'est plus à couvrir.

Le troisième problème, spécifique aux classifieurs à base de règles d'association, est qu'il existe une grande variété de mesures pour ces règles : la fréquence et la confiance, mais également le  $\chi^2$ , l'intensité d'implication, le lift, *etc.* [Tan *et al.* (2002); Vaillant *et al.* (2005)]. Toutes ne donnent pas les mêmes performances [Lan *et al.* (2005)].

## 2.4 État de l'art de la classification à l'aide d'associations

Historiquement, la première méthode de classification à l'aide de règles d'association est CBA [Liu *et al.* (1998)]. Les règles sont pondérées par leur confiance et seule la meilleure est conservée. La méthode de référence CMAR [Li *et al.* (2001)] utilise plusieurs règles pour classer une même instance, et les règles sont pondérées par un  $\chi^2$  normalisé entre la prémisse et la conclusion, puis sélectionnées par couverture de l'échantillon d'apprentissage. L3 [Baralis & Garza (2002)] étend le procédé de couverture, en utilisant les règles qui en sont normalement exclues. L3m [Baralis & Garza (2003)] améliore L3 en utilisant de multiples règles, comme CMAR.

Les tendances actuelles visent à une sélection précise des règles utiles pour les rendre accessibles à l'expert [Zaïane & Antonie (2005)]. Des procédés d'optimisation permettent également d'apprendre comment choisir les règles utiles [Zaïane & Antonie (2005)]. Enfin, le principe de couverture, dépendant de l'ordre des objets, est contesté dans [Wang & Karypis (2006)] et des optimisations pour les grandes bases  $y$  sont présentées. En effet, l'extraction de règles d'association est coûteuse et doit profiter des techniques récentes d'extraction de règles non redondantes [Pasquier *et al.* (2005); Zaki (2000)]. [Bouzouita & Elloumi (2007)] fournit une première contribution dans ce sens.

La comparaison des performances entre les différentes approches est complexe pour plusieurs raisons. D'une part, les prototypes ne sont pas tous disponibles et les expériences correspondantes ne sont pas reproductibles. D'autre part, les articles comparent leur performances à celles de classifieurs « universels », comme C4.5 [Quinlan (1993)], Foil [Quinlan & Cameron-Jones (1993)], Ripper [Cohen (1995)] ou CPAR [Yin & Han (2003)] mais pas toujours à d'autres propositions utilisant les règles d'association. Les comparaisons s'effectuent donc traditionnellement sur les chiffres fournis dans les articles de référence.

## 2.5 Règles d'associations généralisées

Nous nous intéressons à la généralisation des règles d'association pour lesquelles la conclusion et certains attributs de la prémisse peuvent être négatifs. Selon l'angle de

la logique, cela revient à voir une règle classique  $X \rightarrow Y$  comme la conjonction des règles  $X \rightarrow y$  pour  $y \in Y$ . Pour les disjonctions, la forme  $X \rightarrow y_1 \vee \dots \vee y_k$  étend la forme  $X \rightarrow y$ . D'un point de vue logique un attribut positif dans une disjonction en conclusion revient au même que cet attribut, négatif, en prémisse : on a  $(X \rightarrow y \vee Y) \equiv (X \wedge \bar{y} \rightarrow Y)$ , et dualement  $(X \wedge x \rightarrow Y) \equiv (X \rightarrow \bar{x} \vee Y)$ .

Ainsi, pour introduire la négation en prémisse comme en conclusion, nous définissons les règles généralisées comme suit. Cette vision nous sera utile pour l'extraction. Notons que ces règles sont appelées *règles disjonctives* dans [Calders & Goethals (2003)].

### Définition 1 (règle d'association généralisée)

Une règle d'association généralisée basée sur  $Z = X \cup Y$  est une expression de la forme  $X \rightarrow \forall Y$ , où  $X$  et  $Y$  sont deux motifs classiques disjoints. Sa fréquence est  $\mathcal{F}(X)$ , sa profondeur est le nombre d'attributs de  $Y$ . Elle est exacte si chaque objet contenant tous les attributs de  $X$  contient au moins un attribut de  $Y$ .

On remarque que la fréquence d'une règle généralisée est celle de sa prémisse, et non celle de l'union de sa prémisse et de sa conclusion, comme pour les règles classiques. En effet, dans la règle  $X \rightarrow \forall Y$ , la conclusion est une disjonction et cela aurait peu de sens de raisonner sur la fréquence de  $X \cup Y$ . Notons cependant que dans le cas où  $Y$  contient un seul attribut, la règle généralisée  $X \rightarrow \forall Y$  est équivalente à la règle classique *exacte* (de confiance 1)  $X \rightarrow Y$  et leurs fréquences concordent.

D'autre part, la notion de confiance n'est pas utilisée pour les règles généralisées, car il ne paraît pas judicieux d'introduire de l'incertitude à la fois par une disjonction en conclusion et par une confiance non maximale. Sur notre exemple de la table 1, nous avons la règle exacte  $a_2a_3 \rightarrow a_5 \vee a_6$  de fréquence 3. Cette règle généralisée résume les deux règles classiques  $a_2a_3 \rightarrow a_5$  de fréquence 1 et de confiance 1/3 et  $a_2a_3 \rightarrow a_6$  de fréquence 2 et de confiance 2/3. Enfin, la confiance n'est utilisée dans les méthodes de classifications que pour obtenir des règles non redondantes. Pour la méthode CMAR que nous adaptons ici en généralisant la forme des règles utilisées, c'est une mesure de  $\chi^2$  qui pondère les règles et affine leur sélection. Nous proposons donc une adaptation de cette mesure aux règles généralisées à la section 3.2.

## 3 Classification à base de règles d'association généralisées

Nous présentons maintenant une adaptation de CMAR [Li *et al.* (2001)] avec des règles *généralisées*. Nous montrons d'abord comment extraire un ensemble complet de règles non redondantes, puis comment les valoriser dans un processus de classification.

CMAR est notre méthode de référence car elle est techniquement très aboutie (règles non redondantes, sélection par  $\chi^2$  et couverture, vote de multiples règles) et produit de remarquables performances. En modifiant le type de règles qu'elle utilise, cela permet une meilleure compréhension de leur apport.

### 3.1 Extraction de règles généralisées non redondantes

Construire des classifieurs utilisant des règles d'association nécessite tout d'abord d'extraire les règles pertinentes de la base de données. Nous voyons ces règles sous leur forme généralisée, telle que définie dans le paragraphe précédent. Comme dans l'approche standard, nous nous intéressons alors à des règles *fréquentes*, dont la fréquence dépasse un seuil fixé.

Extraire les règles généralisées est difficile. Dans [Boulicaut *et al.* (2000)], une approche par extraction contrainte de motifs généralisés (contenant des attributs et des négations d'attributs) est proposée pour la découverte de règles avec négations. Concernant la classification supervisée, [Antonie & Zaïane (2004)] construit les règles classiques  $X \rightarrow \bar{Y}$ ,  $\bar{X} \rightarrow Y$  ou  $\bar{X} \rightarrow \bar{Y}$  lorsque la corrélation de  $X \rightarrow Y$  est négative, mais cette approche n'est pas complète. Dans [Baralis & Garza (2006)], les auteurs utilisent une démarche énumérative pour faire apparaître une à une des négations dans les règles. La méthode que nous proposons se différencie des propositions précédentes en utilisant les récents développements des représentations condensées des motifs fréquents, afin d'obtenir un ensemble exhaustif de règles généralisées non redondantes.

#### *Minimalité*

Les procédés de classification classiques (*i.e.* CBA, CMAR) évitent la redondance entre les règles extraites en ne conservant que les règles classiques de meilleure confiance. Lorsque deux règles ont même confiance, on ne conserve que celle de plus grande fréquence. Enfin, lorsque les fréquences coïncident, on conserve les règles qui ont la plus petite prémisse. Ces opérations sont réalisées par filtrage de l'ensemble de règles extraites. Nous proposons ici d'extraire directement les règles généralisées à *prémisse et conclusion minimales* selon l'inclusion des motifs et nous caractérisons formellement cette redondance.

#### **Définition 2 (règle redondante, prémisse et conclusion minimales)**

Une règle généralisée exacte  $X \rightarrow \forall Y$  est dite redondante dans une base de données  $r$  s'il existe une règle exacte  $X' \rightarrow \forall Y$  telle que  $X' \subset X$  et  $\mathcal{F}(X') = \mathcal{F}(X)$ , ou s'il existe une règle exacte  $X \rightarrow \forall Y'$  telle que  $Y' \subset Y$ . Si une règle n'est pas redondante dans  $r$ , elle est dite à prémisse et conclusion minimales dans  $r$ .

Cette définition caractérise une forme de redondance que l'on pourrait qualifier de syntaxique. Une règle sera en effet redondante s'il existe une règle équivalente qui s'écrit plus simplement. Dans notre exemple de la table 1, la règle  $a_7 \rightarrow a_4$  est exacte, donc  $a_1 a_7 \rightarrow a_4$  est redondante, tout comme  $a_7 \rightarrow a_2 \vee a_4$ . Il s'agit d'une première étape de sélection des règles, effectuée directement pendant l'extraction selon la méthode qui suit.

#### *Extraction des prémisses et conclusions*

Une règle classique  $X \rightarrow Y$  est à prémisse minimale dans  $r$  si  $X$  a la propriété d'être un motif *libre, clé* ou *générateur minimal* [Calders *et al.* (2006)] dans  $r$ , c'est-à-dire s'il n'existe aucune règle classique exacte de la forme  $X_1 \rightarrow X_2$  avec  $X_1 \cup X_2 = X$ . Nous

calculons donc les règles généralisées non redondantes en extrayant d’abord les motifs libres, qui forment les prémisses potentielles des règles non redondantes.

Pour cela, on observe que la contrainte de liberté sur les prémisses est *anti-monotone*, c’est-à-dire que si un motif est libre, alors tous ses sous-ensembles le sont également. On peut donc extraire les motifs libres (de fréquence supérieure à un certain seuil) en utilisant un algorithme par niveaux classique [Mannila & Toivonen (1997)]. Chaque motif libre et fréquent ainsi extrait servira de prémisses à une ou plusieurs règles.

Il s’agit ensuite de calculer les conclusions non redondantes (*i.e.* minimales)  $Y$  telles que  $X \rightarrow \forall Y$  soit exacte, c’est-à-dire de calculer les ensembles minimaux d’attributs ayant un élément en commun avec tous les objets supportant la prémisse. Cela revient à extraire les *traverses minimales* de l’hypergraphe constitué par les objets contenant tous les attributs de  $X$ . Ce calcul est difficile [Fredman & Kachiyan (1996)], c’est pourquoi nous limitons les règles extraites à une profondeur (nombre d’attributs en conclusion) bornée par une valeur  $l$ , en limitant la profondeur de l’algorithme décrit dans [Boros *et al.* (2003)]. Dans la pratique, il s’avère difficile de dépasser  $l = 3$  sans nuire définitivement aux performances ou voir apparaître des phénomènes de sur-apprentissage.

Cette méthode d’extraction fonctionne également pour les règles d’association classiques exactes, quand  $l = 1$ . Ces règles sont cependant trop strictes pour effectuer une bonne classification, aussi autorisons-nous des exceptions en nombre limité dans la conclusion [Crémilleux & Boulicaut (2002)], un paramètre que nous notons  $\delta$ .

Autoriser un nombre d’exception pour une règle d’association classique revient dans une certaine mesure à accepter des règles dont la conclusion est une disjonction. Par exemple, sur les données de la table 1, la règle  $a_2 \rightarrow a_3$  est exacte à une exception près et  $a_2 \rightarrow a_3 \vee a_7$  est exacte. Il ne paraît pas judicieux d’autoriser à la fois un nombre limité d’exceptions et une disjonction dans la conclusion. Dans nos expériences,  $\delta$  sera nul si  $l > 1$ .

Pour conclure, l’obtention des règles généralisées non redondantes consiste à calculer les motifs  $X$  libres et fréquents, puis les ensembles minimaux  $Y$  d’au plus  $l$  attributs qui intersectent tous les objets contenant  $X$ . Chaque  $Y$  donne lieu à la règle exacte  $X \rightarrow \forall Y$ , fréquente, à prémisse et conclusion minimales par construction. L’approche garantit en outre que l’on calcule bien toutes les règles qui vérifient ces propriétés.

### 3.2 Utilisation de la négation

Le classifieur que nous proposons, à base de règles d’associations généralisées, fonctionne sur le principe de CMAR [Li *et al.* (2001)]. Toutes les règles d’association généralisées non redondantes sont extraites selon la méthode exposée précédemment. Nous ne gardons que les règles contenant une (et une seule) valeur de classe  $c$ . Si une telle règle est de la forme  $Xc \rightarrow \forall Y$  (valeur de classe en prémisse), alors nous la voyons comme la règle  $X\bar{Y} \rightarrow \bar{c}$ . Une telle règle exclut la classe  $c$  pour les objets contenant tous les attributs de  $X$  et aucun de  $Y$ . Si la règle est de la forme  $X \rightarrow c \vee Y$  (valeur de classe en conclusion), alors nous la voyons comme la règle  $X\bar{Y} \rightarrow c$ .

#### Définition 3 (règles prescrivant/excluant une classe)

Soit  $o \in \mathcal{O}$  un objet, soient  $X, Y \subseteq \mathcal{A}$  des motifs, et soit  $c$  une valeur de classe. Si l’on a  $X \subseteq o$  et  $o \cap Y = \emptyset$ , alors on dit que la règle  $Xc \rightarrow \forall Y$  exclut la classe  $c$  pour  $o$ , et



que la règle  $X \rightarrow c \vee Y$  prescrit la classe  $c$  pour  $o$ .

Ainsi, pour un nouvel objet à classer, un certain nombre de règles prescrivent ou excluent une ou plusieurs classes. Chaque règle contribue positivement ou négativement à la classe selon sa pondération. Dans CMAR, la mesure utilisée est un  $\chi^2$  pondéré, qui quantifie la corrélation entre la prémisse  $X$  et la conclusion  $c$  d'une règle classique. Le calcul est effectué à partir du tableau de contingence de  $X$  et de  $c$ . Pour une règle généralisée  $X\bar{Y} \rightarrow c$ , une adaptation naturelle de ce principe nous conduirait à travailler sur un tableau de contingence entre  $X\bar{Y}$  et  $c$ .

Cependant, nous disposons ici de règles dont la conclusion est une disjonction et il est nécessaire d'adapter cette mesure. Une règle  $X \rightarrow c \vee Y$  prescrivant une classe sera intéressante si  $c$  est majoritairement présent avec  $X$ . Pour quantifier cet intérêt, nous mesurons sur les objets contenant  $X$  le  $\chi^2$  obtenu sur le tableau de contingence entre  $\bar{Y}$  et  $c$ . Nous obtenons ainsi une mesure pertinente du degré d'association entre  $\bar{Y}$  et  $c$  pour les objets contenant  $X$ . Des détails sur la mesure de  $\chi^2$  peuvent être trouvés dans [Brin *et al.* (1997)] et la définition du  $\chi^2$  normalisé figure dans [Li *et al.* (2001)].

Le principe du vote des règles sur les objets de test est simple. Lorsqu'un objet contient la prémisse d'une règle prescrivant une classe, celle-ci reçoit une contribution égale à notre mesure de  $\chi^2$ . Pour une règle excluant une classe, la contribution à cette classe est l'opposé du  $\chi^2$ . La décision finale est prononcée pour la classe ayant reçu la somme de contributions la plus importante.

### 3.3 Résultats

Notre méthode de classification à base d'associations généralisées est appelée *l $\delta$ -miner* :  $l$  désigne la profondeur des règles et  $\delta$  le nombre d'exceptions autorisées. Dans la pratique,  $\delta > 0$  que si  $l = 1$ . Pour évaluer ses performances, nous avons effectué pour chaque base de test de l'UCI [Blake & Merz (1998)] une série de 18 expériences avec un support minimal de 1, 2 ou 5%, un  $\delta$  de 0, 1, 2, ou 3, et une profondeur  $l$  variant de 1 à 3. Lorsque  $l \geq 2$ , le mécanisme d'exceptions est désactivé ( $\delta = 0$ ), et inversement lorsque  $\delta \geq 1$ , la longueur de conclusion est celle des règles classiques ( $l = 1$ ). Les paramètres de *l $\delta$ -miner* indiqués dans le tableau 2 sont ceux qui ont donné le meilleur score.

Ce tableau indique les scores de classification en 10-cross-validation sur une large variété de benchmarks, pour les méthodes de référence C4.5, CBA et CMAR, et précise les résultats pour *l $\delta$ -miner* et ses paramètres. Les colonnes indiquent :

dataset :	le nom du jeu de données de l'UCI
cl :	le nombre de classes
obj :	le nombre d'objets
attr :	le nombre d'attributs
c45 :	le score de C4.5, repris de [Li <i>et al.</i> (2001)]
cba :	le score de CBA, fourni dans [Liu <i>et al.</i> (1998)]
cmar :	le score de CMAR, indiqué dans [Li <i>et al.</i> (2001)]
<i>l<math>\delta</math>-miner</i> :	le score de notre méthode

dataset	caractéristiques			méthodes référence			$l\delta$ - miner	paramètres				temps (sec.)
	cl	obj	attr	c45	cba	cmar		type	$l$	$\delta$	$\gamma$	
anneal	6	898	73	94.80	<b>97.90</b>	97.30	93.50	=	3	0	1	331
austral	2	690	55	84.70	84.90	86.10	<b>87.50</b>	+	1	1	2	65
auto	5	202	137	80.10	78.30	78.10	<b>81.20</b>	+	1	0	1	399
breast	2	699	26	95.00	96.30	<b>96.40</b>	94.70	+	1	3	1	10
cleve	2 <sup>a</sup>	303	43	78.20	82.80	82.20	<b>83.80</b>	+	1	2	5	6
crx	2	690	59	84.90	84.70	84.90	<b>86.50</b>	+	1	1	1	189
german	2	1000	76	72.30	73.40	<b>74.90</b>	74.60	+	2	0	1	812
glass	6 <sup>b</sup>	214	34	68.70	<b>73.90</b>	70.10	66.30	+	2	0	5	9
heart	2	270	38	80.80	81.90	82.20	<b>84.00</b>	+	1	3	2	16
hepatitis	2	155	45	80.60	81.80	80.50	<b>82.50</b>	+	1	2	2	24
horse	2	368	75	82.60	82.10	82.60	<b>83.70</b>	+	1	0	5	16
hypo	2	3163	47	<b>99.20</b>	98.90	98.40	94.70	+	1	1	2	3627
iono	2	351	100	90.00	92.30	91.50	<b>92.60</b>	=	1	1	5	360
iris	3	150	15	<b>95.30</b>	94.70	94.00	95.10	+	2	0	5	2
lymph	4	148	63	73.50	77.80	83.10	<b>83.70</b>	+	2	0	1	18
pima	2	768	26	<b>75.50</b>	72.90	75.10	74.30	+	1	2	1	12
sonar	2	208	234	70.20	77.50	79.40	<b>80.70</b>	=	1	1	5	485
tic-tac	2	958	29	99.40	99.60	99.20	<b>100.0</b>	+	1	1	5	2
vehicle	4	846	58	<b>72.60</b>	68.70	68.80	67.50	=	1	0	1	930
wine	3	178	45	92.70	<b>95.00</b>	<b>95.00</b>	94.50	+	1	1	5	4
zoo	7	101	43	92.20	96.80	97.10	<b>98.10</b>	=	1	0	2	14
moyenne				83.97	85.34	85.57	85.72					

TAB. 2 – Récapitulatif des scores de classification (en gras le score de la meilleure méthode).

<sup>a</sup> cinq classes dans les données originales mais les quatre dernières sont fusionnées.

<sup>b</sup> sept classes déclarées mais la classe 4 est absente

Les dernières colonnes indiquent les paramètres de  $l\delta$ -miner pour le meilleur score sur la série d'expériences.

type : le type de règles utilisées pour classer : on note + lorsque seules les règles prescrivant une classe sont utilisées, - pour les règles excluant une classe (ce type n'apparaît pas dans le tableau car elles ne fournissent jamais le meilleur score), = pour l'union de ces deux types de règles

$l$  : la longueur des conclusions de règles

$\delta$  : le nombre maximal d'exceptions tolérées

$\gamma$  : le support minimal d'extraction (en %)

Enfin, la dernière colonne indique le temps nécessaire à la réalisation complète de l'expérience en 10-cross-validation. Ce temps fournit une indication de la difficulté de la tâche, qui réside essentiellement dans le calcul des règles.

Ce tableau montre que notre adaptation  $l\delta$ -miner de CMAR est opérationnelle et atteint des performances comparables à CMAR. Dans certains cas, nous arrivons même à

améliorer les scores. La section suivante détaille ces résultats en fonction des différents types de règles.

## 4 Discussion

L'apport de la négation pour la classification supervisée est discuté selon deux axes.

1. Quel est l'intérêt d'utiliser des règles excluant des classes ? C'est le cas pour les règles contenant une valeur de classe en prémisse.
2. La présence des négations dans la prémisse est-elle bénéfique ? La variation de la profondeur de la règle (longueur de la conclusion) permet d'examiner l'apport des négations.

### 4.1 Impact de l'exclusion de la classe

La figure 1 permet de comparer la performance des règles prescrivant une classe (noté  $l\delta\text{-miner+}$ ), l'excluant ( $l\delta\text{-miner-}$ ) ou les deux types de règles ( $l\delta\text{-miner=}$ ). Cette figure indique les meilleurs scores de chaque méthode, pour la série de 18 expériences pour chaque base donnée en prémisse. La référence est le meilleur score de  $l\delta\text{-miner}$ , reporté à la table 2.

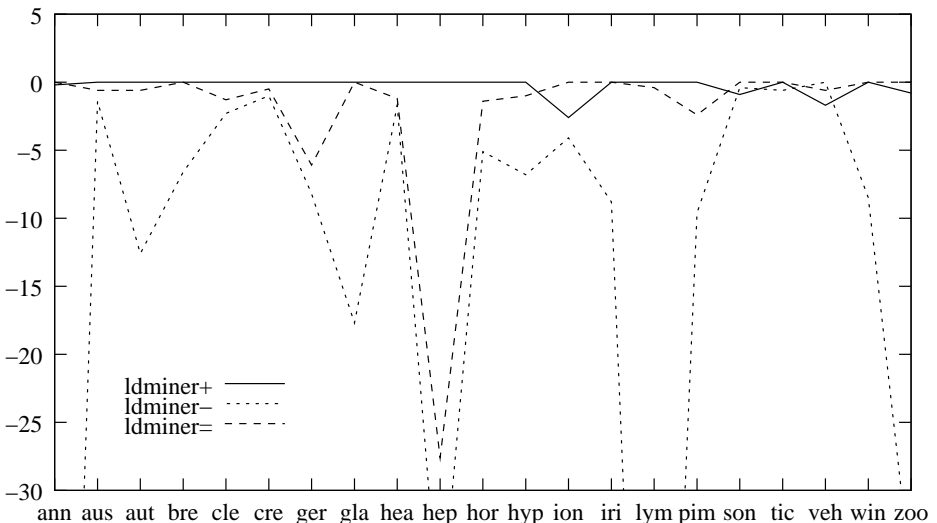


FIG. 1 – Impact de l'exclusion de la classe sur le score de classification. La référence est le meilleur score de la méthode sur tous les paramètres testés (cf. table 2).

On constate que l'apport des règles excluant les classes est marginal. D'une part elles ne peuvent se suffire à elles-mêmes pour construire un classifieur, d'autre part leur combinaison avec les règles prescrivant une classe n'améliore légèrement les performances

que pour 5 bases sur 21 : anneal (0.2%), iono (2.6%), sonar (0.9%), vehicule (1.1%), zoo (0.8%) sur 21.

Rappelons que les règles excluant les classes ne sont pas plus difficiles à extraire que les autres. Si cela permet d'améliorer le score sur certaines bases, il serait dommage de s'en priver. Il suffit en effet de considérer les règles qui comportent une valeur de classe en prémisses sans besoin d'adapter l'extracteur.

## 4.2 Impact des négations en prémisses

La figure 2 montre l'impact de la longueur de la conclusion pour les règles généralisées que nous utilisons, c'est-à-dire du nombre de négations en prémisses. Les scores indiqués sont relatifs au meilleur score de  $l\delta$ -miner, figurant dans la table 2. Nous ne reportons ici que des scores obtenus avec des règles prescrivant une classe, car les expérimentations précédentes ont montré que l'apport des règles excluant une classe est minime. Le support minimum d'extraction est de 1%. Lorsque  $l = 1$ , il n'y a pas de négation en prémisses. Quand  $l = 2$  ou  $l = 3$ , il y a une ou deux négations en prémisses.

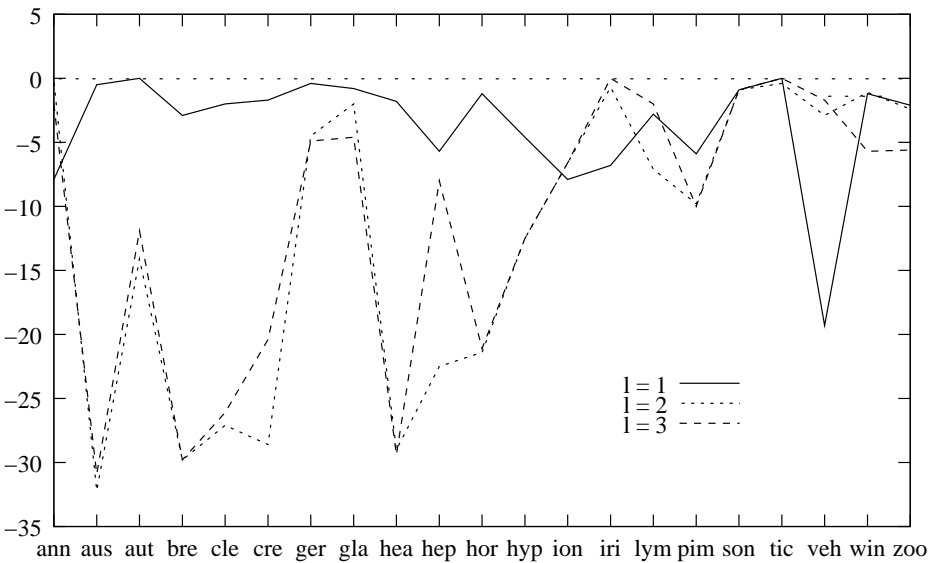


FIG. 2 – Impact de la longueur de la conclusion sur le score de classification. La référence est le meilleur score de la méthode sur tous les paramètres testés (cf. table 2).

Lorsque  $l = 1$ , on retrouve le cas des règles d'association classiques exactes, et ces règles fournissent très souvent le meilleur score. Ce score est cependant souvent en-dessous du score de référence, obtenu en autorisant un petit nombre d'exceptions aux règles (*i.e.*  $\delta = 1, 2, 3$ ) ou en augmentant le support minimal jusqu'à 5%.

Lorsque la longueur de la conclusion augmente à 2 ou 3, on obtient des règles moins fortes et de façon équivalente, on trouve un ou deux littéraux négatifs dans la prémisses. Leur impact est là encore marginal, bien que plus spectaculaire que celui des exclusions

de classe. Il concerne 3 bases de test : aneel (7,7%), iris (6,8%), vehicule (17,6%) et améliore sensiblement les performances.

Selon notre intuition, augmenter la longueur de la conclusion permet de découvrir plus de règles généralisées. Cette méthode peut s'avérer avantageuse lorsque les données sont faiblement corrélées et recèlent peu d'associations classiques. Dans cette configuration, il est difficile de trouver des règles de classification et les performances du classifieur sont mauvaises. Or, plus la longueur autorisée de la conclusion est grande, plus on obtient de règles : les performances en classification sont dans ce cas meilleures qu'avec des règles classiques. Cela a été mis à profit dans [Vernier *et al.* (2007)] pour améliorer la classification de textes d'opinions.

Il semble illusoire d'espérer améliorer les performances avec des conclusions encore plus longues, d'autant que le surcoût de l'extraction est très pénalisant. De plus, ces règles sont souvent trop spécifiques et tendent vers un apprentissage *par cœur* des données.

## 5 Conclusion - perspectives

En suivant le modèle de CMAR, reconnu en classification à base d'associations, nous avons construit un classifieur avec des règles d'association généralisées, pouvant contenir des littéraux négatifs en prémisses et prescrire une classe ou l'exclure (comportant un littéral négatif en conclusion). Sur 21 bases de test de l'UCI, l'impact de ces négations est marginal et peu spectaculaire. L'utilisation de simples règles avec exceptions suffit à obtenir la meilleure performance dans la grande majorité des cas.

L'utilisation de règles généralisées en classification est donc délicate. Les règles excluant les classes sont aisées à obtenir et peuvent apporter une légère plus-value. Les règles généralisées de profondeur supérieure à 1 sont nettement plus difficiles à extraire que les règles classiques. La plus-value obtenue, qui se traduit par la présence de négations en prémisses, n'apporte le raffinement escompté que pour les données peu corrélées, comme indiqué à la section 4.2.

## Références

- AGRAWAL R. & SRIKANT R. (1994). Fast algorithms for mining association rules. In *Intl. Conference on Very Large Data Bases (VLDB'94), Santiago de Chile, Chile*, p. 487–499.
- ANTONIE M.-L. & ZAÏANE O. (2004). An associative classifier based on positive and negative rules. In *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'04), Paris, France*.
- BARALIS E. & GARZA P. (2002). A lazy approach to pruning classification rules. In *IEEE International Conference on Data Mining, ICDM 02 Maebashi City, Japan*.
- BARALIS E. & GARZA P. (2003). Majority classification by means of association rules. In *PKDD*, p. 35–46.
- BARALIS E. & GARZA P. (2006). Associative text categorization exploiting negated words. In *SAC '06 : Proceedings of the 2006 ACM symposium on Applied computing*, p. 530–535, New York, NY, USA : ACM.

- BLAKE C. & MERZ C. (1998). UCI repository of machine learning databases.
- BOROS E., ELBASSIONI K., GURVICH V. & KHACHIAN L. (2003). An efficient implementation of a quasi-polynomial algorithm for generating hypergraph transversals. In *11th Annual European Symposium on Algorithms (ESA'03), Budapest, Hungary*.
- BOULICAUT J.-F., BYKOWSKI A. & JEUDY B. (2000). Towards the tractable discovery of association rules with negations. In *Fourth Int. Conference on Flexible Query Answering Systems FQAS'00*, p. 425–434.
- BOUZOUITA I. & ELLOUMI S. (2007). Integrated generic association rule based classifier. In *DEXA '07 : Proceedings of the 18th International Conference on Database and Expert Systems Applications (DEXA 2007)*, p. 514–518, Washington, DC, USA : IEEE Computer Society.
- BRIN S., MOTWANI R. & SILVERSTEIN C. (1997). Beyond market baskets : generalizing association rules to correlations. In *SIGMOD 1997*, p. 265–276.
- CALDERS T. & GOETHALS B. (2003). Minimal k-free representations of frequent sets. In *Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03), Cavtat-Dubrovnik, Croatia*, p. 71–82.
- CALDERS T., RIGOTTI C. & BOULICAUT J.-F. A. (2006). *A Survey on Condensed Representations for Frequent Sets*, In *Constraint-based mining and Inductive Databases*, p. 64–80. LNCS 3848. Springer.
- COHEN W. W. (1995). Fast effective rule induction. In *ICML*, p. 115–123.
- CRÉMILLEUX B. & BOULICAUT J. (2002). Simplest rules characterizing classes generated by delta-free sets. In SPRINGER, Ed., *International Conference on Knowledge Based Systems and Applied Artificial Intelligence (Expert System)*, Cambridge, UK, p. 33–46.
- FREDMAN M. & KACHIAN L. (1996). On the complexity of dualization of monotone disjunctive normal forms. *Journal of Algorithms*, **21**(2), 618–628.
- LAN Y., JANSSENS D., WETS G. & CHEN G. (2005). Improving associative classification by incorporating novel interestingness measures. In *ICEBE '05 : Proceedings of the IEEE International Conference on e-Business Engineering*, p. 282–288, Washington, DC, USA : IEEE Computer Society.
- LI W., HAN J. & PEI J. (2001). Cmar : Accurate and efficient classification based on multiple class-association rules. In *IEEE International Conference on Data Mining (ICDM'01), San Jose, USA*.
- LIU B., HSU W. & MA Y. (1998). Integrating classification and association rules mining. In *International Conference on Knowledge Discovery and Data Mining (KDD'98), New York, USA*, p. 80–86.
- MANNILA H. & TOIVONEN H. (1997). Levelwise search and borders of theories in knowledge discovery. *Data Mining and Knowledge Discovery*, **1**(3), 241–258.
- PASQUIER N., TAOUIL R., BASTIDE Y., STUMME G. & LAKHAL L. (2005). Generating a condensed representation for association rules. *Journal Intelligent Information Systems (JIIS)*, **24**(1), 29–60.
- QUINLAN J. R. (1993). *C4.5 : Programs for Machine Learning*. Morgan Kaufmann Publishers.
- QUINLAN J. R. & CAMERON-JONES R. M. (1993). Foil : a midterm report. In *European Conference on Machine Learning*, p. 3–20.

- TAN P.-N., KUMAR V. & SRIVASTAVA J. (2002). Selecting the right interestingness measure for association patterns. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26, 2002, Edmonton, Alberta, Canada*, p. 32–41.
- VAILLANT B., MEYER P., PRUDHOMME E., LALLICH S., LENCA P. & BIGARET S. (2005). Mesurer l'intérêt des règles d'association. In *EGC*, p. 421–426.
- VERNIER M., MATHET Y., RIOULT F., CHARNOIS T., FERRARI S. & LEGALLOIS D. (2007). Classification de textes d'opinions : une approche mixte n-grammes et sémantique. In *3ème Défi Fouille de Texte 2007*.
- WANG J. & KARYPIS G. (2006). On mining instance-centric classification rules. *IEEE Trans. Knowl. Data Eng.*, **18**(11), 1497–1511.
- YIN X. & HAN J. (2003). Cpar : Classification based on predictive association rules. In *proceedings of the 2003 SIAM Int. Conf. on Data Mining (SDM'03)*, San Fransisco, CA.
- ZAIANE O. R. & ANTONIE M.-L. (2005). On pruning and tuning rules for associative classifiers. In *Knowledge-Based Intelligent Information and Engineering Systems, 9th International Conference, KES 2005, Melbourne, Australia, September 14-16, 2005, Proceedings, Part III*, p. 966–973.
- ZAKI M. (2000). Generating non-redundant association rules. In *ACM SIGKDD international conference on Knowledge discovery and data mining, Boston, USA*, p. 34–43.