

Qualité des règles et des opérateurs en découverte de connaissances floues

Maurice Bernadet

IRIN / Ecole Polytechnique de l'Université de Nantes
Rue Christian Pauc, La Chantrerie
BP 60601, 44306 Nantes Cedex 3, France
Maurice.Bernadet@polytech.univ-nantes.fr

Résumé. Nous décrivons ici l'évaluation de la qualité des règles et celle de la qualité des opérateurs dans un système de découverte de connaissances floues. Nous indiquons d'abord comment il est possible de définir des partitions floues pour remplacer les attributs classiques par des attributs flous. Nous généralisons ensuite à des règles floues trois indices de qualité utilisés pour des règles classiques et nous détaillons un algorithme d'extraction de règles floues. Etant donné que les logiques floues fournissent une infinité d'opérateurs logiques, nous proposons une méthode pour évaluer la qualité d'un jeu d'opérateurs flous quand les règles extraites doivent être exploitées dans des systèmes à base de connaissances à l'aide du Modus Ponens Généralisé. Nous appliquons cette évaluation sur un exemple simple et nous exposons quelques résultats de ces méthodes sur des bases de données réelles. Nous indiquons les jeux d'opérateurs flous qui s'avèrent de meilleure qualité et en quoi ces résultats correspondent à la théorie. En rappelant des résultats sur la réduction de règles floues, nous constatons que les jeux d'opérateurs dont la qualité est la meilleure pour utiliser le Modus Ponens Généralisé dans des systèmes à base de connaissances, ne sont pas les mieux adaptés à la réduction des règles.

1. Introduction

Quand on réalise un système de découverte de connaissances, des choix initiaux doivent être effectués : le processus peut être supervisé ou non et l'on peut représenter les connaissances à l'aide d'arbres de décision, de règles d'associations, de réseaux neuronaux... Pour le système que nous avons conçu, nous avons choisi, pour des raisons de simplicité, un processus supervisé et une représentation des connaissances utilisant des règles. Ensuite, considérant qu'il est souvent difficile d'interpréter des règles sur des intervalles d'attributs numériques et que des seuils stricts provoquent souvent des coupures trop brutales entre intervalles, nous avons décidé d'utiliser une représentation des connaissances basées sur les logiques floues. Les logiques floues peuvent être considérées comme des généralisations des logiques à valeurs multiples et autorisent des valeurs de vérité intermédiaires entre le vrai et faux (Zadeh 1974). Elles permettent d'exprimer les connaissances de manière plus naturelle que les logiques booléennes classiques, en utilisant des attributs gradués dans des formules comme "X est plutôt haut" (ou "X est haut est plutôt vrai"), "Y est un peu chaud" (ou "X est chaud est un peu vrai")... En outre, les logiques floues offrent le choix de nombreux opérateurs logiques (Klir et Yuan 1995), ce qui permet une bonne diversité dans l'expression des connaissances.

Qualité des règles et des opérateurs en découverte de connaissances floues

L'utilisation des logiques floues demande en premier lieu de convertir des attributs classiques dans une représentation floue ; pour cela, il est nécessaire de définir pour chaque attribut classique une correspondance entre ses valeurs possibles et les valeurs de vérité de chacun des attributs flous associés. C'est le rôle du partitionnement flou, qui permet ensuite de traduire les valeurs des attributs classiques dans les valeurs des attributs flous correspondants.

Après cette opération, appelée aussi "fuzzification", il est possible d'employer un algorithme d'extraction de connaissances construit sur les mêmes mécanismes que ceux utilisant la logique classique. On peut remarquer ici que les logiques floues disposent de méthodes spécifiques d'extraction de connaissances, utilisant en particulier des algorithmes génétiques, pour rechercher un ensemble de poids d'attributs dans les règles floues afin de représenter au mieux les données ; nous ne considérons pas ici cette possibilité : nos méthodes sont une généralisation aux connaissances floues des méthodes "classiques".

Notre algorithme d'extraction de connaissances effectue une recherche exploratoire dans l'arbre des règles possibles en évaluant la qualité de chaque règle pour décider si elle doit être gardée ou non ; cette qualité est évaluée par trois indices : la confiance en la règle, son support et un indice moins habituel appelé l'intensité d'implication. Après un rappel des principes de ces indices, nous indiquons comment nous les avons généralisés aux connaissances floues et nous décrivons notre algorithme d'extraction de règles floues.

Les logiques floues offrent de nombreuses définitions des opérateurs logiques, en particulier l'implication. Si l'on veut seulement que les règles découvertes puissent être analysées par un expert humain, la nature de l'implication importe peu, mais, si les règles découvertes doivent être traitées par un système à base de connaissances à l'aide du Modus Ponens Généralisé, un choix judicieux des opérateurs flous est important. Pour trouver un jeu d'opérateurs flous bien adapté dans ce cas, nous avons limité le nombre d'opérateurs utilisables et nous avons défini une méthode pour déterminer la qualité d'un jeu d'opérateurs flous compte tenu de la base de données considérée. Après un exemple simplifié sur lequel nous mettons en oeuvre ces méthodes, nous présentons quelques résultats sur des bases de données plus conséquentes, puis nous indiquons les jeux d'opérateurs dont la qualité apparaît la meilleure et la conformité de ces résultats avec la théorie. Nous constatons ensuite que ces jeux d'opérateurs de meilleure qualité pour l'utilisation du Modus Ponens Généralisé, ne sont pas les mieux adaptés à la réduction d'ensembles de règles floues.

Nous concluons sur les intérêts et les inconvénients de l'extraction de connaissances floues et nous envisageons les perspectives de notre système.

2. Le processus de partitionnement flou ("fuzzification")

2.1 Définition de partitions floues

Rappelons que les logiques floues évaluent la vérité d'une proposition floue "*X est A*" comme le degré auquel l'entité *X* appartient au sous-ensemble flou *A* :

$$\text{Vérité}(\text{"X est A"}) = \mu_A(X),$$

$\mu_A(X)$ étant la fonction d'appartenance (ou fonction caractéristique) du sous-ensemble flou *A*.

Les (sous-)ensembles flous permettent de définir des "pseudo-partitions floues" ou C-partitions, dans lesquelles chaque valeur d'un attribut peut être classée dans plusieurs classes floues, avec un total des degrés d'appartenance égal à 1. Ces pseudo-partitions floues permettent la conversion d'attributs en attributs flous, donnant alors la valeur de vérité des propositions floues associées. Pour un attribut continu C_A , variant de $\min C_A$ à $\max C_A$, on peut définir une pseudo-partition floue de plusieurs manières (Bezdek et Harris 1978), (Lesmo et al. 1988). Nous n'évoquons pas ici le partitionnement flou des attributs non numériques, mais toutes les méthodes décrites ci-après s'appliquent aussi à ces attributs.

La méthode la plus simple divise l'intervalle $[\min C_A, \max C_A]$ en n sous-intervalles, avec un petit pourcentage de recouvrement entre deux intervalles adjacents et donne à chaque sous-intervalle un nom symbolique lié à sa position. Par exemple, on peut diviser l'intervalle $[\min C_A, \max C_A]$ en 5 sous-intervalles avec un chevauchement d'environ 20 %, donnant alors 5 attributs flous : très négatif, plutôt négatif, moyen, plutôt positif et très positif (Figure 1).

Les classes floues peuvent aussi être définies par les experts et l'on peut proposer 3 ou 5 classes comme des options standard. Des nombres différents de classes peuvent aussi être employés, mais un trop grand nombre de classes risque de ralentir fortement le processus de découverte de connaissances.

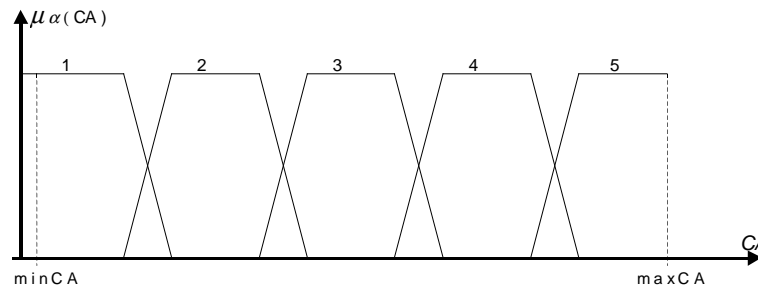


FIG. 1 - Une pseudo-partition floue ($\alpha=1$ très négatif ; $\alpha=2$ plutôt négatif ; $\alpha=3$ moyen ; $\alpha=4$ plutôt positif ; $\alpha=5$ très positif).

D'autres méthodes permettent d'extraire le nombre de classes et de définir les classes floues à partir des données. On considère pour cela les valeurs des attributs donnant la même conclusion et, si possible, on regroupe ces valeurs dans les mêmes ensembles flous, avec une valeur d'appartenance égale au taux d'échantillons donnant cette conclusion. Ces méthodes emploient souvent les histogrammes des valeurs des attributs associés à chaque conclusion. Il est également possible de concevoir une méthode plus satisfaisante en généralisant aux logiques floues des méthodes de discrétisation optimales telles celles étudiées dans (Zighed et al. 1999). Ainsi, nous utilisons, depuis peu, des méthodes de "clusterisation" que nous avons adaptées aux sous-ensembles flous.

2.2 "Fuzzification" d'une base de données

Une fois que les classes floues ont été définies pour chaque attribut, on peut convertir la valeur associée pour chaque item, c'est à dire un enregistrement (ou une "ligne") de la base de données, en faisant correspondre à cette valeur les degrés d'appartenance de l'attribut

classique à chacune des classes : le degré d'appartenance $\mu_i(V)$ de d'un attribut classique de valeur V à une classe floue i est celui qui correspond à V dans la classe i (Figure 2).

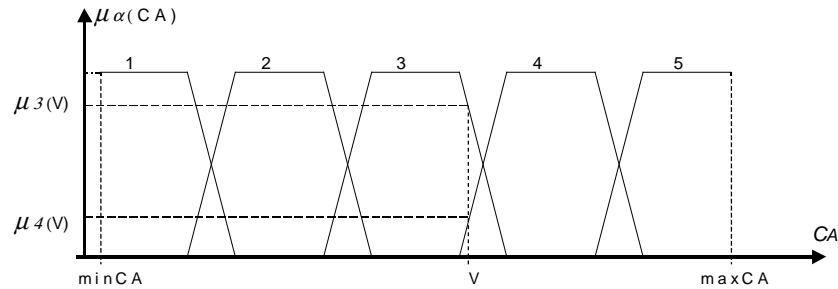


FIG. 2 - Évaluation des degrés d'appartenance d'une valeur V de l'attribut continu C_A aux sous-classes floues (ici, seuls μ_3 et μ_4 ne sont pas nuls)

3. Évaluation de la qualité de règles floues

3.1 Indices de qualité des règles classiques

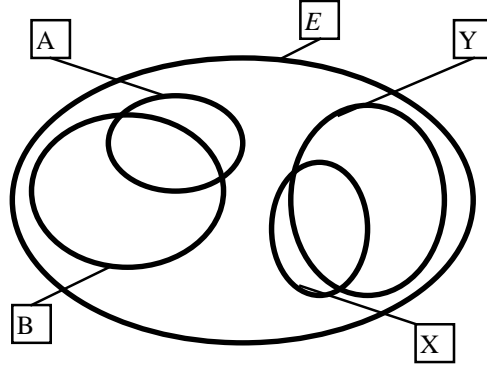
Plusieurs indices peuvent être employés pour évaluer la qualité des règles classiques (Agrawal et al. 1993) ; nous avons choisi trois d'entre eux : la confiance, le support et un indice moins courant, l'intensité d'implication.

- La *confiance* d'une règle "si a alors b " exprime la probabilité conditionnelle de b quand a est vrai ; elle peut être évaluée par $n_{a \wedge b} / n_a$, en appelant $n_{a \wedge b}$ le nombre d'articles (ou "items") vérifiant " a et b " et n_a le nombre d'articles vérifiant a .

- Le *support* peut être défini comme le taux d'occurrences des articles vérifiant " a et b " par rapport à tous les articles de la base de données ; ainsi, le support de la règle s'évalue comme $n_{a \wedge b} / nE$, $n_{a \wedge b}$ étant le nombre d'articles vérifiant " a et b " et nE le nombre total d'articles dans la base de données.

- L'*intensité d'implication* est un indice exprimant la qualité d'une règle. Cet indice, défini par (Gras et Larher 1992), est basé sur des concepts simples de probabilités : puisque les cardinalités de deux sous-ensembles A et B sont déterminées par les objets de la base de données appartenant à A et B , nous considérons deux sous-ensembles aléatoires X et Y ayant respectivement mêmes cardinalités que A et B . L'implication $a \Rightarrow b$ est caractérisée par la relation $A \subset B$ et ses contre-exemples sont associés au sous-ensemble $A \cap \overline{B}$. Nous comparons la cardinalité de $A \cap \overline{B}$ (donnée par la base de données) avec la variable aléatoire

donnée par la cardinalité de $X \cap \overline{Y}$, en supposant qu'il n'y a aucune lien entre X et Y (Figure 3). Si la cardinalité de $A \cap \overline{B}$ est exceptionnellement petite comparée à l'espérance de la distribution sur les cardinalités de $X \cap \overline{Y}$, nous acceptons "si a alors b " comme étant une règle.

FIG. 3 - X et Y varient aléatoirement dans E .

L'intensité de l'implication $a \Rightarrow b$ peut donc être définie comme le complément à un de la probabilité pour la variable aléatoire "cardinalité de $X \cap \bar{Y}$ " d'être plus petite que la cardinalité de $A \cap \bar{B}$, fournie par l'échantillon. Elle est définie par

$$\varphi(a, \bar{b}) = 1 - P[\text{Card}(X \cap \bar{Y}) \leq \text{Card}(A \cap \bar{B})] ;$$

Si l'on appelle $n = \text{Card}(E)$, $n_a = \text{Card}(A)$, $n_{\bar{a}} = \text{Card}(\bar{A})$, $n_b = \text{Card}(B)$, $n_{\bar{b}} = \text{Card}(\bar{B})$, $n_{a \cap b} = \text{Card}(A \cap B)$, $n_{a \cap \bar{b}} = \text{Card}(A \cap \bar{B})$, la variable aléatoire $\text{Card}(X \cap \bar{Y})$ vérifie une distribution hyper géométrique (Fleury 1995) :

$$P[\text{Card}(A \cap \bar{Y}) = k] = \frac{C_{n_a}^k \cdot C_{n - n_a}^{n - n_b - k}}{C_n^{n - n_b}} = \frac{C_{n_a}^k \cdot C_{n_{\bar{a}}}^{n_{\bar{b}} - k}}{C_n^{n_{\bar{b}}}}$$

et

$$P[\text{Card}(X \cap \bar{Y}) \leq \text{Card}(A \cap \bar{B})] = \sum_{\substack{i=0 \\ i \geq n_a - n_b}}^{\text{Card}(A \cap \bar{B})} \frac{C_{n_a}^i \cdot C_{n_{\bar{a}}}^{n_{\bar{b}} - i}}{C_n^{n_{\bar{b}}}} .$$

L'intensité d'implication présente plusieurs propriétés intéressantes (Briand 1995). D'abord, sa valeur augmente avec la taille de l'ensemble d'apprentissage, alors que les autres indices restent constants. De plus, la prise en compte de quelques contre-exemples de plus change peu une valeur d'intensité d'implication forte, mais progressivement le doute vient et de nouveaux contre-exemples finissent par en causer la chute. Cet indice est également bien adapté aux données bruitées, puisqu'un petit nombre de contre-exemples ne provoque pas le rejet de l'implication ; il évite aussi de considérer des règles $a \Rightarrow b$ quand la proposition b est vraie pour presque tous les exemples de l'ensemble d'étude (il n'est pas surprenant alors que presque tous les exemples où a est vraie soient des exemples où b est vraie).

3.2 Adaptation des indices classiques aux règles floues

Une *règle floue* peut-être considérée comme la généralisation d'une règle classique : elle associe dans la partie prémisses une conjonction de propositions floues avec, en conclusion, une proposition floue. Les propositions utilisées prennent en compte les attributs flous d'un même item (ou article) d'une base de données. Si X_1, X_2, \dots, X_n et Y sont des attributs flous, pouvant se voir associer des modalités respectives A_1, A_2, \dots, A_n et B , une règle floue est de la forme

Si " X_1 est A_1 " et " X_2 est A_2 " ... et " X_n est A_n " alors " Y est B ",

ou, plus formellement

$"X_1 \text{ est } A_1" \wedge "X_2 \text{ est } A_2" \dots \wedge "X_n \text{ est } A_n" \rightarrow "Y \text{ est } B"$,

où A_1, A_2, \dots, A_n sont des sous-ensembles flous et non des sous-ensembles classiques.

Par exemple, on peut écrire des règles comme

"Si la température est élevée alors la pression est forte",

"Si la température est basse et l'humidité est forte alors la saturation est proche"...

Deux catégories d'indices permettent d'évaluer des règles floues : des indices basés exclusivement sur la théorie des ensembles flous (Aguilar Martin et Lopez de Mantara 1982), (Lesmo et al. 1988) ou des indices développés en logique classique et généralisés aux connaissances floues (Rives 1990), (Weber 1992), (Zeidler et Schlosser 1995). Nous avons choisi cette seconde catégorie, et généralisé les trois indices d'évaluation classiques en considérant que le nombre d'éléments qui satisfont une proposition p associée à un ensemble flou P de fonction d'appartenance μ_P est la cardinalité précise de l'ensemble flou P , telle qu'elle est définie par (Zadeh 1968) : $Card(P) = \sum_{x \in E} \mu_P(x)$.

Cette notion de cardinalité a été récemment critiquée : dans le cas où une forte proportion d'items possède de faibles degrés d'appartenance aux sous-ensembles flous considérés, la comparaison entre ces cardinalités peut conduire à des résultats aberrants (Delgado et al. 2000). Ce problème ne devrait pas pouvoir se produire dans le cadre de notre réalisation, étant donné que les sous-ensembles flous que nous utilisons constituent une pseudo-partition, construite soit à l'aide d'un expert, soit par une méthode de clusterisation et que cette construction fournit des sous-ensembles flous dont le noyau (ensemble des éléments dont le degré d'appartenance au sous-ensemble flou est de 1) couvre une grande partie du support (ensemble des éléments dont le degré d'appartenance au sous-ensemble flou est supérieur à 0). Pour prévenir le risque d'un mauvais partitionnement, nous vérifions que le support des sous-ensembles flous n'est pas supérieur à un seuil σ de pourcentage de leur noyau. C'est à dire que, si nous notons $Support(A) = \{x | \mu_A(x) \geq 0\}$ et $Noyau(A) = \{x | \mu_A(x) = 1\}$, nous vérifions que $(Card(Support(A)) - Card(Noyau(A))) / Card(Noyau(A)) \leq \sigma$. Dans le cas contraire, nous utilisons comme cardinalité d'un sous-ensemble flou A sur un univers de discours U , une formule inspirée de (Ralescu 1995) et (Wygralak 1999), en ne prenant en compte que les degrés d'appartenance supérieurs ou égaux à 0.5 (ce que l'on appelle aussi l' α -coupure de niveau 0.5) :

$$Card(A) = \sum_{x \in E, \mu_A(x) \geq 0.5} \mu_A(x)$$

Cependant, si l'on considère une implication $a \Rightarrow b$ entre deux propositions floues a et b , associées à des sous-ensembles flous A et B de fonctions d'appartenance μ_A et μ_B , une conjonction floue T (norme triangulaire ou t-norme) et un complément flou $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$, on peut en général écrire (Bernadet et al. 1996) :

$$\begin{aligned} n_A &= \text{Card}(A) = \sum_{x \in E} \mu_A(x), & n_{\bar{A}} &= \text{Card}(\bar{A}) = \sum_{x \in E} \mu_{\bar{A}}(x) = \sum_{x \in E} (1 - \mu_A(x)), \\ n_{\bar{B}} &= \text{Card}(\bar{B}) = \sum_{x \in E} \mu_{\bar{B}}(x) = \sum_{x \in E} (1 - \mu_B(x)), \\ n_{A \cap B} &= \text{Card}(A \cap B) = \sum_{x \in E} \mu_{A \cap B}(x) = \sum_{x \in E} T(\mu_A(x), \mu_B(x)), \\ n_{A \cap \bar{B}} &= \text{Card}(A \cap \bar{B}) = \sum_{x \in E} \mu_{A \cap \bar{B}}(x) = \sum_{x \in E} T(\mu_A(x), (1 - \mu_B(x))) . \end{aligned}$$

La confiance d'une règle, son support et son intensité d'implication sont alors exprimés par les mêmes formules que ci-dessus, mais en utilisant les cardinalités des sous-ensembles flous au lieu des cardinalités des ensembles classiques.

4. Algorithme d'extraction de connaissances

Nos méthodes sont l'adaptation aux logiques floues de méthodes classiques (c'est à dire non floues) d'extraction de connaissances. Nous devons ici remarquer que les logiques floues disposent de méthodes qui leur sont propres pour extraire des ensembles de règles. Ces méthodes spécifiques, plus globales, peuvent utiliser des réseaux neuro-mimétiques (Buckley et Hayashi 1995), (Halgamuge et Glesner 1994), (Heinz 1995),... ou des algorithmes génétiques (Bonarini 1996), (Murata et al. 1998), (Roubos et al. 2001), (Surmann 2000),... afin de construire des ensembles de règles reflétant de manière quasi-optimale les données. Nous avons préféré adapter les méthodes classiques d'extraction de connaissances aux connaissances floues, parce que ces méthodes, plus analytiques, permettent le suivi des mécanismes au cours de leur mise en œuvre.

Notre algorithme, décrit formellement dans l'annexe 1, emploie une stratégie en profondeur d'abord, pour évaluer toutes les règles qui peuvent être construites à partir d'un ensemble de propositions. Pour réduire le nombre de règles à évaluer, nous limitons le nombre de propositions dans la partie prémisse d'une règle ; ainsi, nous employons 4 seuils $\alpha, \beta, \gamma, \delta$ et une règle est retenue si sa confiance C est plus grande que α , son support S plus grand que β , son intensité d'implication I supérieur à γ et si la longueur L de sa prémisse est d'au plus δ propositions.

Cet algorithme explore, sans le mémoriser, l'arbre des règles que l'on peut construire à partir des propositions floues associées aux attributs autres que celui cherché en conclusion (notre système étant supervisé, l'attribut à utiliser comme conclusion est fixé par l'utilisateur). Cette exploration utilise les attributs dans l'ordre où ils sont implantés dans la base de données (l'ordre des "colonnes") ; nous n'avons pas cherché, comme le fait par exemple (Quinlan 1993) à utiliser en premier les attributs les plus discriminants, pour ne pas accroître

Qualité des règles et des opérateurs en découverte de connaissances floues

la complexité de nos algorithmes, et pour tenir compte du fait que les critères de qualité retenus permettent un filtrage au niveau des règles.

L'algorithme que nous utilisons explore donc l'arbre des règles en partant de sa racine (prémisse vide), et utilise pour cela deux procédures de balayage :

"forward" ajoute, si possible, une proposition floue non encore utilisée à ce niveau, à la partie prémisse de la règle,

"backward" retire la dernière proposition floue de la partie prémisse (la plus "à droite"), et la remplace par sa suivante éventuelle ; si cette proposition n'a pas de suivante (dernière modalité du dernier attribut), la proposition qui se trouve alors la plus à droite est retirée pour, si possible, être remplacée et ainsi de suite. Quand il ne reste plus de proposition en prémisse, l'arbre des prémisses a été complètement exploré.

Avec, par exemple 3 attributs flous {a, b, c} possédant chacun trois modalités : faible, moyen, fort, l'arbre des règles sera exploré en considérant successivement les prémisses dans l'ordre de la table 1 :

1°) a=faible a=faible \wedge b=faible a=faible \wedge b=faible \wedge c=faible a=faible \wedge b=faible \wedge c=moyen a=faible \wedge b=faible \wedge c=fort a=faible \wedge b=moyen a=faible \wedge b=moyen \wedge c=faible a=faible \wedge b=moyen \wedge c=moyen a=faible \wedge b=moyen \wedge c=fort a=faible \wedge b=fort a=faible \wedge b=fort \wedge c=faible a=faible \wedge b=fort \wedge c=moyen a=faible \wedge b=fort \wedge c=fort	2°) a=fort a=fort \wedge b=faible a=fort \wedge b=faible \wedge c=faible a=fort \wedge b=faible \wedge c=moyen a=fort \wedge b=faible \wedge c=fort a=fort \wedge b=moyen a=fort \wedge b=moyen \wedge c=faible a=fort \wedge b=moyen \wedge c=moyen a=fort \wedge b=moyen \wedge c=fort a=fort \wedge b=fort a=fort \wedge b=fort \wedge c=faible a=fort \wedge b=fort \wedge c=moyen a=fort \wedge b=fort \wedge c=fort	3°) a=fort a=fort \wedge b=faible a=fort \wedge b=faible \wedge c=faible a=fort \wedge b=faible \wedge c=moyen a=fort \wedge b=faible \wedge c=fort a=fort \wedge b=moyen a=fort \wedge b=moyen \wedge c=faible a=fort \wedge b=moyen \wedge c=moyen a=fort \wedge b=moyen \wedge c=fort a=fort \wedge b=fort a=fort \wedge b=fort \wedge c=faible a=fort \wedge b=fort \wedge c=moyen a=fort \wedge b=fort \wedge c=fort
4°) b=faible b=faible \wedge c=faible b=faible \wedge c=moyen b=faible \wedge c=fort	5°) b=moyen b=moyen \wedge c=faible b=moyen \wedge c=moyen b=moyen \wedge c=fort	6°) b=fort b=fort \wedge c=faible b=fort \wedge c=moyen b=fort \wedge c=fort
7°) c=faible	8°) c=moyen	9°) c=fort

TAB. 1 – *Ordre d'examen des prémisses dans l'exploration de l'arbre des règles*

5. Qualité des opérateurs flous

L'algorithme d'extraction de connaissances met en évidence un ensemble de règles intéressantes, mais il est nécessaire de disposer d'un autre mécanisme pour évaluer les implications floues. Il faut alors prendre en compte le fait que les opérateurs flous acceptent beaucoup de définitions possibles (Klir et Yuan 1995).

5.1 Principaux opérateurs flous

Le complément flou (négation floue) est généralement réalisé à l'aide du complément standard $C(a) = 1 - a$; c'est le seul que nous considérons.

Une conjonction floue ("et" flou) peut être définie par une t-norme (norme triangulaire) T qui est une fonction de $[0,1] \times [0,1]$ dans $[0,1]$, caractérisée par :

$$\begin{aligned} T(0, 0) &= T(0, 1) = T(1, 0) = 0, \\ T(1, 1) &= 1, \\ T(a, b) &= T(b, a) \text{ (commutativité)}, \\ T(a, T(b, c)) &= T(T(a, b), c) \text{ (associativité)}, \\ \forall a, a', b, b' \quad a < a' \text{ et } b < b' &\Rightarrow T(a, b) \leq T(a', b') \text{ (monotonie)}. \end{aligned}$$

Les quatre premiers axiomes assurent que l'intersection floue garde les propriétés d'une intersection classique pour des ensembles classiques. D'autres axiomes sont souvent ajoutés :
la continuité de $T(x, y)$,
la sous-idempotence : $T(a, a) \leq a$.

Une conjonction floue peut donc être choisie parmi de nombreuses classes de t-normes ; on utilise souvent :

le minimum (Zadeh)	$T(a, b) = \min(a, b)$,
l'intersection probabiliste	$T(a, b) = a * b$,
la différence bornée ("bold" ou de Lukasiewicz)	$T(a, b) = \max(0, a + b - 1)$.

Une disjonction floue ("ou" flou) peut être définie par une t-conorme (conorme triangulaire) \perp qui est une fonction de $[0,1] \times [0,1]$ dans $[0,1]$, caractérisée par :

$$\begin{aligned} \perp(1, 1) &= \perp(0, 1) = \perp(1, 0) = 1, \\ \perp(0, 0) &= 0, \\ \perp(a, b) &= \perp(b, a) \text{ (commutativité)}, \\ \perp(a, \perp(b, c)) &= \perp(\perp(a, b), c) \text{ (associativité)}, \\ \forall a, a', b, b' \quad a < a' \text{ et } b < b' &\Rightarrow \perp(a, b) \leq \perp(a', b') \text{ (monotonie)}. \end{aligned}$$

Les quatre premiers axiomes assurent que l'union floue garde les propriétés d'une union classique pour des ensembles classiques. D'autres axiomes sont souvent ajoutés :

la continuité de $\perp(x, y)$,
la sur-idempotence : $\perp(a, a) \geq a$.

Le choix d'une disjonction floue peut être effectué parmi de nombreuses classes de t-conormes ; on utilise souvent :

Qualité des règles et des opérateurs en découverte de connaissances floues

l'union standard (Zadeh) :	$\perp(a,b) = \max(a,b)$,
l'union probabiliste :	$\perp(a,b) = a+b-a*b$,
la somme bornée (Lukasiewicz) :	$\perp(a,b) = \min(1, a+b)$.

L'intersection et l'union classiques sont liées par les deux lois de de Morgan : $\neg(A \cap B) = \neg A \cup \neg B$ et $\neg(A \cup B) = \neg A \cap \neg B$; il est souhaitable que ces relations soient aussi vérifiées pour des ensembles flous. Dans ce cas on a : $C(T(a,b)) = \perp(C(a), C(b))$ et $C(\perp(a,b)) = T(C(a), C(b))$; l'on dit alors que la t-norme T et la t-conorme \perp sont duales par rapport au complément flou. Parmi les couples de t-norme et t-conorme duales par rapport au complément flou standard $C(a) = 1-a$, on peut citer :

- le minimum et le maximum : $T(a,b) = \min(a,b)$, $\perp(a,b) = \max(a,b)$;
- le produit et la somme probabilistes : $T(a,b) = a*b$, $\perp(a,b) = a+b-a*b$;
- la différence et la somme bornées : $T(a,b) = \max(0, a+b-1)$, $\perp(a,b) = \min(1, a+b)$.

L'implication floue est une fonction I de $[0,1] \times [0,1] \rightarrow [0,1]$ qui pour toutes les valeurs de vérité a et b de deux propositions floues respectivement p et q , définit la valeur de vérité $I(a,b)$ de la proposition "Si p alors q ". I peut être définie sous différentes formes, non équivalentes en logique floue, mais qui le deviennent en logique classique.

1) En logique classique, on peut définir I par :

$$\forall a \in \{0,1\} \quad \forall b \in \{0,1\} \quad I(a,b) = \neg a \vee b, \quad (1)$$

ce qui donne en logique floue :

$$\forall a \in [0,1] \quad \forall b \in [0,1] \quad I(a,b) = \perp(C(a), b); \quad (2)$$

2) la logique classique permet aussi d'écrire :

$$\forall a \in \{0,1\} \quad \forall b \in \{0,1\} \quad I(a,b) = \max\{x \in \{0,1\} \mid a \wedge x \leq b\}, \quad (3)$$

d'où en logique floue :

$$\forall a \in [0,1] \quad \forall b \in [0,1] \quad I(a,b) = \sup\{x \in [0,1] \mid T(a, x) \leq b\}. \quad (4)$$

Les relations (1) et (3) sont équivalentes, mais les relations (2) et (4) ne le sont pas car elles portent sur des implications floues. La formule (1) peut aussi être écrite sous les formes suivantes :

$$I(a,b) = (\neg a) \vee (a \wedge b) \quad (\text{car } \neg a \vee a = I), \quad (5)$$

$$I(a,b) = (\neg a \wedge \neg b) \vee b, \quad (6)$$

ce qui donne en logique floue :

$$I(a,b) = \perp(C(a), T(a,b)), \quad (7)$$

$$I(a,b) = \perp(T(C(a), C(b)), b) \quad (8)$$

où T , \perp et C doivent satisfaire la loi de Morgan.

Les formules (2), (4), (7) et (8) permettent de construire différentes classes d'implications floues :

Les S-implications sont définies à partir de la formule (2) qui définit une implication floue I à partir d'une t-conorme \perp : $I(a,b) = \perp(C(a), b)$. On définit ainsi à partir des unions floues de base (ou des intersections duales) :

- pour l'union standard, le maximum (intersection duale : le minimum)
l'implication de Kleene-Dienes : $\forall a \in [0,1] \forall b \in [0,1] I(a,b) = \max(1-a,b)$
- pour l'union probabiliste (intersection duale : le produit) ;
l'implication de Reichenbach : $\forall a \in [0,1] \forall b \in [0,1] I(a,b) = 1-a+a*b$
- pour la somme bornée (intersection duale : la différence bornée)
l'implication de Lukasiewicz : $\forall a \in [0,1] \forall b \in [0,1] I(a,b) = \min(1,1-a+b)$.

Les *R-implications* sont définies par la formule (4) qui définit une implication à partir d'une t-norme: $I(a,b) = \sup\{x \in [0,1] \mid T(a, x) \leq b\}$; cela permet de définir :

- pour l'intersection standard (le minimum)
l'implication de Gödel : $\forall a \in [0,1] \forall b \in [0,1] \quad I(a,b) = 1 \text{ si } a \leq b$
 $\quad \quad \quad = b \text{ si } a > b.$
- pour l'intersection probabiliste,
l'implication de Goguen : $\forall a \in [0,1] \forall b \in [0,1] \quad I(a,b) = 1 \text{ si } a \leq b$
 $\quad \quad \quad = b/a \text{ si } a > b.$
- pour la différence bornée,
l'implication de Lukasiewicz : $\forall a \in [0,1] \forall b \in [0,1] \quad I(a,b) = \min(1,1-a+b).$

La classe des *QL-implications*, utilise la relation (7) avec une t-norme T et une t-conorme \perp duales par rapport au complément flou C ; on peut alors définir :

- pour l'union standard, le maximum (intersection duale : le minimum)
l'implication de Zadeh : $\forall a \in [0,1] \forall b \in [0,1] \quad I(a, b) = \max(1-a, \min(a, b)) ,$
- pour l'union probabiliste (intersection duale : le produit) ;
l'implication : $\forall a \in [0,1] \forall b \in [0,1] \quad I(a, b) = 1-a+a^2 * b ,$
- pour la somme bornée (intersection duale : la différence bornée)
l'implication de Kleene-Dienes : $\forall a \in [0,1] \forall b \in [0,1] \quad I(a, b) = \max(1-a, b) .$

Cette classe d'implications s'étant avérée peu intéressante dans le cadre de nos études, nous ne la détaillerons pas plus.

5.2 Méthode d'évaluation de la qualité de jeux d'opérateurs flous

Quand les règles extraites devront être employées dans un système à base de connaissances par application du modus ponens généralisé (M.P.G.), les opérateurs utilisés pour l'extraction de connaissances devront être les mêmes que ceux qui seront alors utilisés. Rappelons que le M.P.G. est le schéma d'inférence suivant :

Qualité des règles et des opérateurs en découverte de connaissances floues

$$\frac{\begin{array}{l} \text{Si } X \text{ est } A \text{ alors } Y \text{ est } B \\ X \text{ est } A' \end{array}}{Y \text{ est } B'}$$

Pour la t-norme T et l'implication I , on a $\mu_{b'}(y) = \sup_{x \in A'} T(\mu_{a'}(x), I(\mu_a(x), \mu_b(y)))$.

Une étude comparative des opérateurs d'implication floue (Kerre 1992) a montré que le modus ponens généralisé donne de meilleurs résultats avec quatre combinaisons : l'implication de Lukasiewicz et l'intersection "bold", l'implication de Kleene-Dienes et le minimum, l'implication de Kleene-Dienes et l'intersection "bold", l'implication de Gödel-Brouwer et l'intersection "bold". Nous avons donc d'abord limité nos essais aux combinaisons de ces opérateurs ; ensuite, nous avons ajouté l'implication de Goguen qui est associée à l'intersection probabiliste.

L'algorithme qui évalue la qualité d'un jeu d'opérateurs flous pour chaque règle mise en évidence par l'extraction de connaissances est donné en annexe 2 ; plus de détails sont donnés dans (Bernadet 1998). Chaque règle peut être composée d'une conjonction de propositions floues en prémisse et d'une proposition floue en conclusion, et l'algorithme utilise un échantillon aléatoire de la base de données pour calculer le nombre d'exemples et de contre exemples pour chaque couple (x, y) de l'implication, puis il évalue le taux de bons exemples, dont la moyenne pondérée par le nombre d'individus associés à (x, y) constitue le critère de qualité du jeu d'opérateurs testé. Si ce critère satisfait l'expert, l'ensemble d'opérateurs flous est gardé ; sinon, un autre ensemble d'opérateurs doit être essayé.

Nos algorithmes utilisent un *opérateur d'agrégation* ; pour disposer d'une évaluation moyenne de l'implication et d'un mécanisme permettant d'exclure les items aberrants, nous avons choisi la *moyenne arithmétique*, qui permet d'utiliser l'écart-type.

6. Exemple simple

Ces algorithmes ont été mis en oeuvre sur des bases de données de référence, et ont généralement donné de bons résultats ; cependant il est parfois arrivé qu'aucune des implications floues considérées ne se soit avérée pleinement adéquate. Dans ces cas, nous proposons d'utiliser une évaluation statistique des conclusions à partir des valeurs de vérité effectives des conséquents (Bernadet 2000). Pour illustrer notre démarche, nous avons préféré présenter ici un exemple simple, plus "parlant" ; nous commenterons ensuite des résultats obtenus sur des bases de données plus conséquentes.

Considérons deux attributs "Taille" et "Pointure" dans un ensemble d'individus qui peuvent être regroupés en deux classes floues : les petits avec des petits pieds, et les grands avec des grands pieds. A ces classes correspondent les règles :

"petite taille" \rightarrow "petite pointure"

"grande taille" \rightarrow "grande pointure"

Dans un premier jeu d'essai (*Figure 4*), les items qui n'appartiennent à aucune classe correspondent à 5% de bruit ajouté.

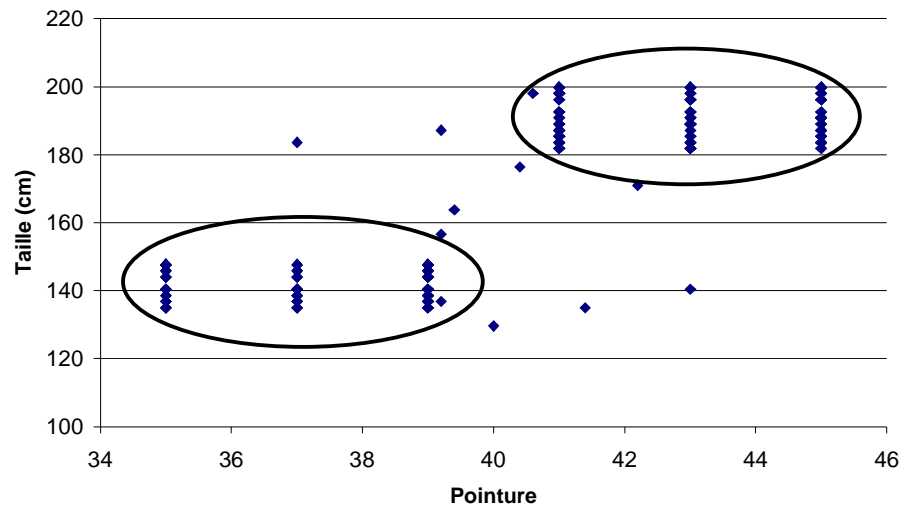


FIG. 4. Répartition des données dans le premier jeu d'essai

Dans un deuxième jeu d'essai (*Figure 5*), les données bruitées représentent 37% de l'échantillon.

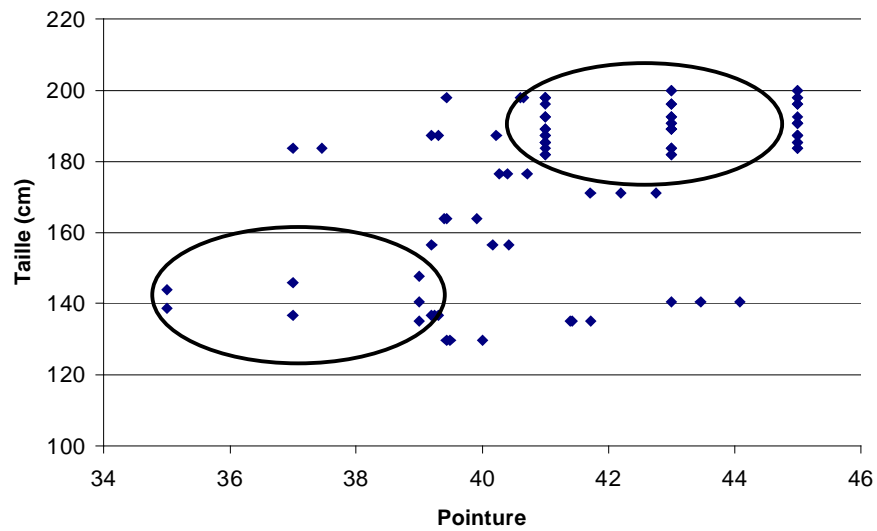


FIG. 5. Répartition des données dans le deuxième jeu d'essai

6.1 Premier jeu d'essai

A) Règle : "taille=petit" \rightarrow "pointure=petit"

Les confiances calculées à l'aide des différentes t-normes sont proches (environ 85% : 0.866 pour l'intersection "bold", 0.85 pour l'intersection probabiliste et 0.844 pour celle de Zadeh). Cette règle est donc intéressante à étudier car ces confiances indiquent qu'environ 85% des personnes de petite taille ont une petite pointure. La table 2 représente la qualité des jeux d'opérateurs pour l'utilisation du Modus Ponens Généralisé sur cette règle, pour chaque implication et chaque t-norme. Les résultats sont très proches les uns des autres quelle que soit la t-norme ou l'implication choisie.

Conjonction Implication	"bold"	de Zadeh	probabiliste
Gödel-Brouwer	0,895923	0,900154	0,897859
Goguen	0,899513	0,899546	0,900154
probabiliste	0,892563	0,896382	0,893538
Kleene-Dienes	0,887846	0,89701	0,890855
Lukasiewicz	0,900154	0,896241	0,897908

TAB. 2 – Qualité des opérateurs pour la règle "taille=petit" \rightarrow "pointure=petit"

B) Règle : "taille=grand" \rightarrow "pointure=petit"

La confiance associée aux différentes t-normes est de 0.148 pour l'intersection "bold", 0.164 pour l'intersection probabiliste et 0.17 pour celle de Zadeh. Cette règle n'est donc pas intéressante puisqu'elle n'est vérifiée que pour environ 16% des personnes grandes.

C) Règle : "taille=grand" \rightarrow "pointure=grand"

La confiance calculée à l'aide des différentes t-normes est de 0.78 pour l'intersection "bold", 0.76 pour l'intersection probabiliste et 0.754 pour celle de Zadeh. Cet échantillon est donc très représentatif avec une probabilité conditionnelle proche de 76,5%.

La table 3 montre des différences assez significatives (1 à 5%) sur l'évaluation de la qualité des opérateurs quelle que soit la t-norme. On note que pour la t-norme "bold", trois implications, Lukasiewicz, Goguen et Gödel-Brouwer, donnent de bons résultats. D'autre part, avec la t-norme probabiliste, ce sont les implications de Goguen, de Gödel-Brouwer et de Lukasiewicz qui donnent de bons résultats, assez similaires (qualité d'adéquation au Modus Ponens Généralisé de plus de 92%).

Conjonction Implication	"bold"	Zadeh	probabiliste
Gödel-Brouwer	0,921295	0,926686	0,923805
Goguen	0,92604	0,926031	0,926686
probabiliste	0,894012	0,918646	0,901187
Kleene-Dienes	0,876628	0,919381	0,889767
Lukasiewicz	0,926686	0,918095	0,921885

TAB. 3 - *Qualité des opérateurs pour la règle "taille=grand" → "pointure=grand"*

6.2 Deuxième jeu d'essai

A) Règle : "taille=petit" → "pointure=petit"

La confiance associée aux différentes t-normes est de 0.523 pour l'intersection "bold", 0.51 pour l'intersection probabiliste et 0.516 pour celle de Zadeh, valeurs proches. Cet échantillon est assez critique : avec une confiance voisine de 52% il est suffisamment grand pour être pris en compte, mais les résultats d'application de la règle peuvent être erronés. Cela est confirmé par le support, voisin de 28%.

La table 4 montre une qualité des opérateurs assez faible (moins de 85%) qui s'explique par le bruit inséré dans l'échantillon ; cette table met en évidence les implications les plus pertinentes. A nouveau, certaines implications se regroupent selon les t-normes, particulièrement les implications de Gödel-Brouwer et Goguen d'une part, celle de Lukasiewicz et celle de Goguen d'autre part.

Conjonction Implication	"bold"	Zadeh	probabiliste
Gödel-Brouwer	0,800913	0,848043	0,825563
Goguen	0,834291	0,840908	0,848043
probabiliste	0,805685	0,784657	0,789694
Kleene-Dienes	0,76613	0,797044	0,778592
Lukasiewicz	0,848043	0,781826	0,808287

TAB. 4 - *Qualité des opérateurs pour la règle taille="petit" → pointure="petit"*

B) Règle : "taille=grand" → "pointure=grand"

La confiance associée aux différentes t-normes est 0.726 pour la t-norme "bold", 0.7 pour la t-norme probabiliste et 0.688 pour celle de Zadeh, valeurs proches. Cet échantillon est donc très représentatif avec une confiance voisine de 70%.

Qualité des règles et des opérateurs en découverte de connaissances floues

Conjonction Implication	"bold"	Zadeh	probabiliste
Gödel-Brouwer	0,915288	0,936152	0,925429
Goguen	0,930192	0,930027	0,936152
probabiliste	0,898016	0,900648	0,896631
Kleene-Dienes	0,872485	0,906273	0,882153
Lukasiewicz	0,936121	0,897394	0,91496

TAB. 5 - Qualité des opérateurs pour la règle "taille=grand" → "pointure=grand"

Les résultats de la table 5 sont encore assez rapprochés. On retrouve les regroupements vus sur la table 4, pour une confiance bien plus élevée ce qui conforte ces regroupements.

En ce qui concerne des bases de données plus conséquentes, nous avons expérimenté nos algorithmes sur plusieurs bases de l'UCI (Blake et Merz 1998), en particulier "Wisconsin Breast Cancer Database", "Wine Recognition Database" et "Ionosphere Database". Les résultats mis en évidence sur notre exemple jouet ont été similaires, mais avec des différences d'évaluation de la qualité des opérateurs moins fortes que dans notre deuxième jeu d'essai, pour lequel la proportion de données bruitées a été volontairement forcée. Nous indiquons quelques exemples des résultats obtenus :

Pour la base de données "Wisconsin Breast Cancer Database", avec trois partitions floues sur chaque attribut en partie prémisse, une confiance minimale de 0.8, une intensité d'implication minimale de 0.9, un support de 5% et des prémisses d'au plus 3 propositions, nous obtenons 331 règles ; si l'on pousse la recherche jusqu'à 6 propositions, on obtient alors 814 règles. Le passage à 9 propositions apporte peu de règles supplémentaires (870 en tout).

L'évolution du nombre de règles en fonction du nombre de classes, construites par une répartition égale sur l'intervalle de variation de chaque attribut est donnée par le tableau ci-dessous :

	3 prémisses au +	6 prémisses au +	9 prémisses au +
2 classes	407 règles	908 règles	954 règles
3 classes	331 règles	814 règles	870 règles
4 classes	250 règles	632 règles	678 règles
5 classes	267 règles	464 règles	466 règles
6 classes	253 règles	471 règles	474 règles
7 classes	255 règles	432 règles	434 règles
8 classes	242 règles	410 règles	412 règles
9 classes	395 règles	672 règles	674 règles

L'on constate que le nombre de règles extraites diminue quand le nombre de classes augmente, jusqu'à 8 classes. Cette profusion du nombre de règles avec des petits nombres de classes se compense par l'imprécision des règles : la confiance moyenne des règles avec 2 classes est beaucoup plus faible que celle obtenue avec un plus grand nombre ; ainsi avec 2

classes et au plus 6 prémisses, seules 35 règles (soit 7%) ont une confiance de 1, alors qu'avec 9 classes et 6 prémisses, 388 règles (soit 57%) ont la même confiance.

Le peu de gain apporté par l'augmentation du nombre de prémisses au-delà de 6 s'explique par le fait que l'apport de nouvelles propositions ne fait en général que spécialiser des règles pour lesquelles la confiance n'est pas totale.

Ainsi, la règle *If Clump Thickness="very small" then Class="benign"*, émerge avec une confiance de 0.964, un support de 28%, une qualité des opérateurs de 91,1%, 90,3% ou 89,6% suivant les opérateurs flous choisis. La spécialisation de cette règle par la prise en compte d'attributs supplémentaires augmente la confiance en réduisant le support, jusqu'à atteindre une confiance de 1 ; la règle est alors : *If Clump Thickness="very small" & Single Epithelial Cell Size="very small" & Bare Nuclei="very little" then Class="benign"*. Son support est alors de 26 % et la qualité des opérateurs de 100%.

Pour les deux autres bases de données les résultats sont similaires, mais, compte tenu du plus grand nombre d'attributs, les nombres de règles générées sont beaucoup plus grands : si l'on garde les mêmes seuils avec 3 classes par attribut, on obtient, pour "Wine" 1092 règles d'au plus 3 prémisses et 13470 d'au plus 6 prémisses, et pour "Ionosphere" 13824 règles d'au plus 3 prémisses. Seul un choix de valeurs de seuils plus sévères permet de réduire ces nombres de règles.

7. Synthèse des résultats

7.1 Qualité des opérateurs pour l'utilisation du modus ponens généralisé

On peut constater sur les exemples précédents que, quelle que soit la règle, pour une t-norme donnée, c'est toujours la même implication qui s'avère de meilleure qualité dans l'évaluation du MPG.

Pour la t-norme "bold", on peut grouper l'implication de Goguen et celle de Lukasiewicz, ces deux implications donnant les meilleurs résultats dans l'utilisation du MPG.

Pour la t-norme de Zadeh, on peut classer les implications en deux groupes. Dans le premier, on trouve les implications de Gödel-Brouwer et de Goguen qui donnent les meilleurs résultats pour le MPG. Dans l'autre groupe, les implications de Lukasiewicz, probabiliste et de Kleene-Dienes donnent des résultats moins bons.

Pour la t-norme probabiliste, on peut regrouper les implications de Gödel-Brouwer et de Goguen, comme donnant les meilleurs résultats.

Plus généralement, nous avons constaté que les jeux d'opérateurs de meilleure qualité pour l'utilisation du MPG correspondaient aux associations entre :

- la t-norme de Lukasiewicz (t-norme "bold") et l'implication de Lukasiewicz,
- la t-norme de Gödel (le minimum de Zadeh) et l'implication de Brouwer-Gödel,
- la t-norme probabiliste et l'implication de Goguen.

Qualité des règles et des opérateurs en découverte de connaissances floues

Ainsi, pour la mise en oeuvre du Modus Ponens Généralisé, les jeux d'opérateurs qui apparaissent expérimentalement de meilleure qualité sont ceux qui associent à une t-norme T , la R-implication I qu'elle définit par :

$$I(a, b) = \sup\{x \in [0, 1] \mid T(a, x) \leq b\}.$$

Ce résultat est justifié théoriquement, en particulier par (Hajek 1998) qui montre que l'implication la mieux adaptée au Modus Ponens Généralisé pour une t-norme donnée est le résidu de cette t-norme (la définition d'une R-implication est en effet le résidu de la t-norme associée).

7.2 Qualité des opérateurs pour la réduction d'ensembles de règles floues

Dans certaines applications, les règles ne sont pas extraites pour construire des systèmes à base de connaissances, mais pour donner aux experts humains une vue synthétique d'une base de données ; pour ce contexte, comme le nombre de règles extraites est souvent élevé, nous avons étudié des méthodes pour agréger les règles.

En logique classique, on peut écrire $(a \Rightarrow c) \wedge (b \Rightarrow c) \vdash (a \vee b) \Rightarrow c$, et nous avons voulu pouvoir procéder de même avec des règles floues, sans avoir à réévaluer la règle floue $(a \vee b) \Rightarrow c$. Ainsi, nous avons cherché parmi les jeux d'opérateurs flous étudiés ci-dessus, ceux qui permettent d'écrire la forme condensée $T(I(a, c), I(b, c)) = I(\perp(a, b), c)$, T étant une t-norme, \perp sa t-conorme duale et I une implication floue. Nous avons alors démontré (Bernadet 2000) que l'implication de Kleene-Dienes, associée au *minimum* pour la t-norme et au *maximum* pour la t-conorme est la seule solution parmi les jeux d'opérateurs flous considérés.

De même, pour conserver la réduction classique $(a \Rightarrow b) \wedge (a \Rightarrow c) \vdash a \Rightarrow b \wedge c$ ou, avec les mêmes notations condensées $T(I(a, b), I(a, c)) = I(a, T(b, c))$, nous avons démontré que le même jeu d'opérateurs flous doit être utilisé.

On peut remarquer à ce propos qu'il n'est pas étonnant que l'implication de Kleene-Dienes associée au *minimum* et au *maximum* conserve les réductions classiques

$$(a \Rightarrow c) \wedge (b \Rightarrow c) \vdash (a \vee b) \Rightarrow c \quad \text{et} \quad (a \Rightarrow b) \wedge (a \Rightarrow c) \vdash a \Rightarrow b \wedge c,$$

puisque cette implication est une S-implication définie à partir du *maximum* : $I(a, b) = \perp(C(a), b)$, et que le *minimum* et le *maximum* sont duaux par rapport au complément standard. On remarque cependant que ce jeu d'opérateurs ne fait pas partie des meilleurs pour exploiter les règles découvertes en utilisant le Modus Ponens Généralisé. Ces résultats montrent de plus que des règles floues ne peuvent généralement pas être traitées comme des règles classiques.

8. Conclusion

Nous avons décrit une généralisation à des connaissances floues de mécanismes classiques de découverte de connaissances, et en particulier la notion de qualité de règle. Pour manipuler des connaissances floues, il est d'abord nécessaire de rendre flous ("fuzzifier") les attributs de la base de données ; à cet effet, la construction de pseudo-

partitions floues peut être effectuée par consultation d'un expert humain ou par une méthode de clusterisation floue. Pour évaluer la qualité des règles, nous avons généralisé aux logiques floues trois indices : la confiance, le support et un indice moins courant, l'intensité d'implication. Nous avons ensuite décrit l'algorithme d'extraction de connaissances floues que nous employons.

Etant donné que les logiques floues permettent l'utilisation de nombreux opérateurs, nous avons restreint le choix du jeu d'opérateurs et nous avons développé une méthode pour valider ce choix, en utilisant comme critère de qualité le taux de bons exemples obtenus par application du Modus Ponens Généralisé sur un échantillon extrait de la base de données. Un exemple simplifié a permis d'illustrer ces mécanismes et nous avons indiqué quelques résultats sur des bases de données plus conséquentes.

Nous avons constaté d'une manière générale que les jeux d'opérateurs flous qui s'avèrent de meilleure qualité pour l'utilisation du modus ponens généralisé, font correspondre une t-norme et la R-implication associée :

- la t-norme de Lukasiewicz (t-norme "bold") et l'implication de Lukasiewicz,
- la t-norme de Gödel (le *minimum* de Zadeh) et l'implication de Brouwer-Gödel,
- la t-norme probabiliste et l'implication de Goguen.

Cependant, pour réduire le nombre de règles proposées aux experts humains, nous avons montré que la réduction de règles floues devrait plutôt utiliser comme opérateurs l'implication de Kleene-Dienes avec le *minimum* comme t-norme et le *maximum* comme t-conorme. Aussi, dans les applications où les règles sont extraites pour donner à des experts humains une vue synthétique d'une base de données, c'est plutôt le choix de ce dernier jeu d'opérateurs flous que nous conseillons.

Nous devons aussi remarquer que l'augmentation de la complexité des calculs, induite par l'utilisation de logiques floues, est relativement faible pour l'extraction des règles, puisque au lieu d'incrémenter (de un) les compteurs de bons et mauvais exemples, les logiques floues nécessitent l'addition de degrés d'appartenance. Les opérations de partitionnement flou (fuzzification), le choix des opérateurs flous et les réductions de règles sont plus complexes, mais les avantages des logiques floues peuvent compenser cela : les intervalles sur des attributs continus sont exprimés par des étiquettes floues plus expressives et les effets de seuils abrupts sont évités.

Parmi les perspectives futures, nous avons l'intention d'améliorer la configuration automatique des partitions floues en tenant compte de la distribution des données ; nous avons déjà généralisé quelques méthodes classiques de clusterisation à la prise en compte d'attributs flous, et nous nous proposons de comparer ces méthodes à des méthodes conçues directement pour des attributs flous. Nous devons également approfondir les mécanismes de réduction des règles floues et nous orienter vers la parallélisation des algorithmes d'extraction.

Références

- Agrawal R., Imilienski T. and Swami A. (1993), Mining Association Rules between Sets of Items in Large Databases, Proc. Conf. on Management of Data, (1993), pp. 207-219, New York: ACM Press.
- Aguilar Martin J. and Lopez De Mantaras R. (1982), The Process of Classification and Learning the Meaning of Linguistic Descriptors of Concepts, M.M. Gupta et E. Sanchez (eds.) Approximate reasoning in decision analysis, North Holland, pp. 165-175.
- Bernadet M., Rose G., Briand H. (1996), FIABLE and Fuzzy FIABLE: two learning mechanisms based on a probabilistic evaluation of implications, Conference IPMU'96, Granada (Spain), July 1996, pp. 911-916.
- Bernadet M. (1998), A knowledge discovery mechanism with evaluation of fuzzy implications, Conference IPMU'98, Paris (France), July 1998.
- Bernadet M. (2000), Basis of a Fuzzy Knowledge Discovery System, Conference Pkdd'2000 (Principles of data Mining and Knowledge Discovery), Lyon, France, Sept. 2000 – Publié dans LNAI vol. 1910, pp. 24-33, Springer 2000.
- Bezdek, J.C. et Harris, J.D. (1978), Fuzzy Partitions and Relations: An Axiomatic Basis for Clustering, Fuzzy Sets and Systems, 1, pp. 111-127.
- Blake C.L. et Merz C.J. (1998), UCI Repository of machine learning databases [<http://www.ics.uci.edu/~mllearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science.
- Bonarini A. (1996), Evolutionary Learning of Fuzzy rules: competition and cooperation, in Fuzzy Modelling: Paradigms and Practice, Kluwer Academic Press, pp. 265-84, 1996.
- Briand H., Djeraba C., Fleury L., Masson Y. et Philippe J. (1995), Contribution of the implication intensity in rules evaluations for knowledge discovery in databases, ECML'95, Heraklion, Crete, April 1995.
- Buckley J. J. et Hayashi (1995), Neural networks for fuzzy systems, Fuzzy Sets and Systems 71 (1995) 265-276.
- Delgado M., Sánchez D., Vila M.A. (2000), Fuzzy Cardinality Based Evaluation of Quantified Sentences, Int. Journal of Approximate Reasoning. Vol. 23, pp. 23-66.
- Fleury L., Masson Y. (1995), Intensity of implication: a measurement in learning machine, IEAAIE'95, 8th International Conference on Industrial and Engineering applications of AI and Expert Systems, 6-8 June 1995, Melbourne, Australia.
- Furuhashi T., Miyata Y., Nakaoka K., et Uchikawa Y. (1995), A new approach to genetic based machine learning and an efficient finding of fuzzy rules - proposal of Nagoya approach, LNAI 1011, pp. 173-189. Springer-Verlag.
- Gras R., Larher A. (1992), L'implication statistique, une nouvelle méthode d'analyse de données, Mathématiques, Informatique et Sciences Humaines n°120.
- Hajek P. (1998), Metamathematics of Fuzzy Logic, Kluwer Academic Publisher, 1998.
- Halgamuge S. K. et Glesner M. (1994), Neural networks in designing fuzzy systems for real world applications, Fuzzy Sets and Systems 65, pp. 1-12.
- Heinz A. (1995), Adaptive Fuzzy Neural Trees, Proc. of IDA-95 Symposium (Advances in Intelligent Data Analysis), Baden-Baden, Germany, 17-19 August 1995, pp. 70-74.
- Kerre Etienne E. (1992), A comparative study of the behaviour of some popular fuzzy implications on the generalized modus ponens, Fuzzy Logic for the management of uncertainty, L. Zadeh, J Kacprzyk (eds.), Wiley, pp. 281-295.

- Klir George J, Bo Yuan (1995), *Fuzzy Sets, and Fuzzy Logics - Theory and Applications*, PrenticeHall, Englewood Cliffs, USA.
- Lesmo L., Saitta L., Torasso P. (1998), *Fuzzy production rules: a learning methodology*, P.P. Wang (ed.) *Advances in Fuzzy Sets, Possibility Theory and Applications*, New York, Plenum Press, pp. 181-198.
- Murata T., Ishibuchi H., Nakashima H., Gen M. (1998), *Fuzzy Partition and Input Selection by Genetic Algorithms for Designing Fuzzy Rule-Based Classification Systems*, LNCS Vol. 1447, Springer-Verlag, pp.407-416.
- Quinlan Ross J. (1993), *C4.5: Programs for Machine Learning*, Morgan Kaufmann, 1993.
- Ralescu A. (1995), *Cardinality, quantifiers and the aggregation of fuzzy criteria*, *Fuzzy Sets and Systems* 69 (1995), pp. 355-365.
- Rives J. (1990), *FID3: Fuzzy Induction Decision Tree*, ISUMA'90, Maryland, USA, December 1990, pp. 457-462.
- Roubos J.A., Setnes M. et Abonyi J. (2001), *Learning fuzzy classification rules from data*, in *Developments in Soft Computing*, John R. and Birkenhead R. Editors, pp. 108-115, Springer-Verlag (<http://citeseer.nj.nec.com/roubos00learning.html>).
- Surmann H. (2000), *Learning a fuzzy rule based knowledge representation*, Proc. of 2nd ICSC symposium on neural computation, NC'2000, Berlin, 23-26 May 2000, pp. 349-355.
- Turksen I.B. (1991), *Measurement of membership functions and their acquisition*, *Fuzzy Sets and Systems* 40, pp. 5-38.
- Weber R. (1992), *Fuzzy-ID3: a Class of Methods for Automatic Knowledge Acquisition*, *Proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks*, Iizuka, Japan, pp. 265-268, July 1992.
- Wygralak M. (1999), *Questions of cardinality of finite fuzzy sets*, *Fuzzy Sets and Systems* 102, pp. 185-210.
- Zadeh L.A. (1968), *Probability Measures of Fuzzy Events*, *Journal of Mathematical Analysis and Applications*, Vol. 23, pp. 421-427, 1968
- Zadeh L.A. (1974), *Fuzzy Logic and its application to approximate reasoning*, *Information Processing* 74, pp. 591-594, 1974.
- Zeidler J., Schlosser M. (1995), *Fuzzy Handling of Continuous-Valued Attributes in Decision Trees*, Proc. ECML-95 Mlnet Familiarization Workshop Statistics, Machine Learning and Knowledge Discovery in Databases, Heraklion, Crete, Greece, April 1995, pp. 41-46.
- Zighed D.A., Rabaseda S., Rakotomalala R., Feschet F. (1999), *Discretization methods in supervised learning*, *Encyclopaedia of Computer Science and Technology*, vol. 40, pp. 35-50, Marcel Dekker inc., 1999.

Annexe 1. Algorithme d'extraction de connaissances

Cet algorithme évalue toutes les règles qui peuvent être construites à partir d'un ensemble de propositions, avec les diverses conclusions floues associées à un attribut choisi par l'utilisateur. Il utilise 4 seuils α , β , γ , δ ; une règle est retenue si sa confiance C est plus grande que α , son support S plus grand que β , son intensité d'implication I supérieure à γ et si la longueur L de sa prémisse est d'au plus δ propositions.

Les paramètres d'entrée sont :

$E = \{e_1, e_2, \dots, e_n\}$ l'ensemble d'apprentissage,

$P = \{p_1, p_2, \dots, p_n\}$ l'ensemble de propositions pouvant décrire les items de E ,

P' l'ensemble des propositions associées aux conclusions possibles,

$D = \{a_1, a_2, \dots, a_m\}$ l'ensemble d'attributs possibles dans les propositions,

$F_{decision}$ est la partition floue associée à l'attribut de décision.

Le paramètre de sortie est R , l'ensemble des règles à garder.

L'algorithme peut s'écrire :

Début

(1) $R = \emptyset$

(2) Pour toutes les valeurs $v_i \in F_{decision}$ faire

(3) Soit T , l'arbre des règles construites sur P avec en conclusion la proposition :

$$p_i = \{a_{decision} = v_i\}$$

(4) Soit B , l'ensemble des observations de E avec la proposition p_i vraie

(5) NoeudCourant = Forward(T , R)

(6) Tant que l'arbre T n'a pas été totalement exploré faire

(7) Soit $r : Prémisse \rightarrow p_i$ la règle associée à NoeudCourant

(8) Soit A , l'ensemble des items de E pour qui $Prémisse$ est vraie

(9) Calculer C , S , I à partir des cardinalités des ensembles E , A et B

(10) Soit L , la longueur de $Prémisse$

(11) Si $(C \geq \alpha)$ et $(S \geq \beta)$ et $(I \geq \gamma)$ et $(L \leq \delta)$

(12) Alors $R = R \cup \{r\}$

(13) Fin Si

(14) Si $(S < \beta)$ ou $(L > \delta)$ ou (NoeudCourant terminal)

(15) Alors NoeudCourant = Backward(T , NoeudCourant)

(16) Sinon NoeudCourant = Forward(T , NoeudCourant)

(17) Fin Si

(18) Fin Tant que

(19) Fin Pour

Fin.

Annexe 2. Evaluation de la qualité d'un jeu d'opérateurs flous

Cet algorithme évalue pour une règle donnée la qualité d'un jeu d'opérateurs flous, en comparant, pour chaque item, le résultat du modus ponens généralisé appliqué aux attributs prémisses de la règle, avec la valeur effective de l'attribut en conclusion.

Debut

1) Pour chaque item :

Évaluer la valeur floue de la prémisse en utilisant la t-norme,
Calculer les valeurs de l'implication choisie.

2) Calculer la moyenne $\eta(x, y)$ et l'écart type $\sigma(x, y)$ pour chaque paire (x, y) où
 x est une valeur de vérité de la prémisse,
 y est une valeur de vérité de la conclusion.

3) Pour chaque exemple :

Si la valeur de son implication pour la paire (x, y) est dans l'intervalle de 95% de confiance pour une loi normale, $[\eta(x, y) - 2 * \sigma(x, y), \eta(x, y) + 2 * \sigma(x, y)]$,

le noter comme exemple positif pour (x, y) ;

sinon

le noter comme un exemple négatif pour (x, y) .

4) Les valeurs de l'implication sont alors les moyennes arithmétiques des bons exemples pour chaque couple (x, y) (x pour la prémisse, y pour la conclusion).

5) Evaluer ensuite l'adéquation de l'implication :

Pour chaque bon exemple de l'échantillon,

Appliquer le Modus Ponens Généralisé,

Calculer la distance entre les valeurs de vérité des conclusions déduites et l'observation,

Si cette distance est supérieure à un seuil fixé,

Ajouter l'item à l'ensemble d'items pour qui la règle est inadéquate, $n-(x, y)$

Sinon

Ajouter l'item à l'ensemble de ceux pour qui la règle est correcte, $n+(x, y)$.

6) Les taux de bons exemples sont alors $\rho(x, y) = n+(x, y) / (n+(x, y) + n-(x, y))$

7) La qualité du jeu d'opérateurs est alors la moyenne pondérée des taux de bons exemples, qui est aussi le taux général de bons exemples :

$$q = \sum_x \sum_y n(x, y) \times \rho(x, y) / \sum_x \sum_y n(x, y) = \sum_x \sum_y n+(x, y) / \sum_x \sum_y (n+(x, y) + n-(x, y))$$

Fin.

Summary

We describe here methods to evaluate the quality of rules and the quality of operators in a fuzzy knowledge discovery system. We first indicate how one can define fuzzy partitions to replace classical attributes by fuzzy ones. We give then a generalization to fuzzy rules of three quality indexes used for classical rules and we detail one algorithm to extract fuzzy rules. Since fuzzy logics provide an infinite number of logical operators, we propose a method to evaluate the quality of a set of fuzzy operators when one wants to exploit extracted rules in a knowledge-based system using the Generalize Modus Ponens. We apply this method to a simplified example and summarize results on real databases. We present then sets of fuzzy operators that appear to be the best ones using this evaluation of quality, and we explain how these results agree with the theory of fuzzy operators. Recalling results on fuzzy rules reduction, we note that the sets of fuzzy operators proving the best for extraction of rules to be used in knowledge based system with the Generalized Modus Ponens, are not well suited for reductions of rules.