Extraction de règles graduelles floues renforcées

Bernadette Bouchon-Meunier, Anne Laurent, Marie-Jeanne Lesot

Extraction de règles graduelles floues renforcées

Mining for Strengthened Gradual Rules

Bernadette Bouchon-Meunier[1], Anne Laurent[2] et Marie-Jeanne Lesot[1]

[1] LIP6 - UPMC - CNRS UMR 7606 - 104 av. du Président Kennedy, 75016 Paris
[2] LIRMM-Univ. Montpellier 2 - CNRS UMR 5506 - 161 rue Ada, 34095 Montpellier

Résumé :

Les règles graduelles, de la forme *plus X est A, plus Y est B*, ont été étudiées de manière intensive au cours de ces dernières années et ont notamment servi dans des systèmes de raisonnement flou basés sur des bases de règles graduelles. Nous nous intéressons ici à des règles que nous nommons renforcées, qui enrichissent ce schéma par une clause supplémentaire introduite par l'expression *d'autant plus que* : ces règles sont de la forme *plus X est A, alors plus Y est B, d'autant plus que Z est C*, la prémisse, la conclusion et le renforcement pouvant être généralisés pour contenir plusieurs clauses. De telles règles graduelles floues renforcées ont montré leur grand intérêt dans de nombreux domaines d'application comme par exemple le domaine de la sécurité routière (*plus la vitesse est élevée, plus le danger est grand, d'autant plus que les virages sont serrés*). Nous exposons ici la définition et la sémantique que nous associons à cette forme de règle, ainsi que les moyens de les mettre en évidence.

Mots-clés :

Fouille de données, règles graduelles, règles floues, renforcement.

Abstract:

Gradual rules such as *the more X is A, then the more Y is B* have been intensively studied for the last years, especially for fuzzy gradual rule-based reasoning. In this paper, we consider strengthened rules that contain a special clause reinforcing the rule, such as *the more X is A, then the more Y is B, all the more as Z is C*, where the premise, the conclusion and the strengthening part can be generalised to several clauses. Such strengthened rules have shown their interest for many applications such as road safety (*the higher the speed, the greater the danger, all the more the tighter the bends*). We study here the definition and semantics of such rules, and how they can be mined.

Keywords:

Data Mining, Gradual Rules, Fuzzy Rules, Strengthening.

1 Introduction

Les règles graduelles floues ont été très étudiées dans le cadre des systèmes de recommandation et de contrôle afin de représenter de manière linguistique les corrélations de variations entre éléments numériques, et d'associer des décisions à des situations [DP92, Del99, GDP04]. Ces règles sont de la forme *plus le mur est proche, plus le freinage est fort* : elles considèrent des attributs (par exemple *mur* ou *freinage*) et des descripteurs flous (par exemple *proche* ou *fort*) dont elles lient les variations.

De manière générale, une règle graduelle floue est de la forme *plus* X_1 *est* A_1 *et ... et plus* X_n *est* A_n*, alors plus* Y_1 *est* B_1 *et ... et plus* Y_p *est* B_p. L'expression *plus* X_1 *est* A_1 *et ... et plus* X_n *est* A_n est le motif graduel antécédent, l'expression *plus* Y_1 *est* B_1 *et ... et plus* Y_p *est* B_p est le motif graduel conséquent, un lien de causalité est établi entre l'antécédent et le conséquent.

De telles règles peuvent être associées à deux types de sémantiques : elles peuvent être considérées comme exprimant des contraintes sur chaque donnée individuellement, imposant que les degrés d'appartenance aux modalités présentes dans la règle vérifient une implication floue, modélisée par une r-implication [DP92, GDP04]. Elles peuvent par ailleurs être interprétées comme des tendances globales qui s'appliquent aux données dans leur ensemble, imposant des corrélations aux variations des valeurs des attributs ou des degrés d'appartenance aux modalités [Hül02, BCS+07, DLT09]. Dans cet article, nous nous plaçons dans ce second cas.

Il existe plusieurs méthodes d'extraction automatique de tels motifs graduels flous définies récemment par [Hül02, BCS+07, DLT08, DLT09, LLR09]. Ces méthodes s'appuient notamment sur des algorithmes de fouille de données basés sur des approches par niveau, extrayant les motifs graduels de taille 2 (liant deux attributs seulement), puis de taille 3, etc. Les relations de causalité au sein de ces motifs graduels flous peuvent ensuite être établies pour définir des règles graduelles floues.

Toutefois ces travaux ne considèrent pas la question de l'influence de renforcement que certains attributs peuvent exercer, et qui sont exprimés linguistiquement dans des clauses introduites par l'expression *d'autant plus que*. De telles règles sont de la forme *plus la distance au mur est proche, plus le freinage est fort, d'autant plus que la vitesse est élevée* : dans cet exemple, l'attribut de vitesse renforce la règle et enrichit par une précision supplémentaire la relation établie entre la distance et le freinage.

Dans cet article nous nous intéressons à ces règles, que nous nommons règles graduelles floues renforcées : après avoir rappelé dans la section 2 les principes de l'extraction de règles graduelles en considérant principalement leur variante floue, nous précisons dans la section 3 la définition des règles graduelles floues renforcées et détaillons les notions de confiance et support qui peuvent leur être attachées. Nous présentons enfin un algorithme permettant de les extraire automatiquement.

2 Travaux associés : règles graduelles floues

Différents algorithmes d'extraction automatique de règles graduelles floues définies en termes de co-variation d'attributs ont été proposés dans la littérature, les principales approches sont détaillées dans la section 2.3.

Les données qu'elles considèrent sont définies selon un schéma $(X_1, \ldots, X_m)$ où les X_i sont des attributs définis sur les domaines numériques, ou univers, $dom(X_i)$. Une base de données DB est alors une instance de ce schéma, définie par un ensemble de lignes (m-uplets) de $dom(X_1) \times \ldots \times dom(X_m)$.

Dans le cas de données floues, ce schéma est un peu modifié : les attributs X_i sont des variables linguistiques floues, associées à des modalités (valeurs linguistiques). On peut ainsi avoir X_1 = vitesse, avec l'ensemble de modalités $\{$*faible, normale, élevée*$\}$. Les données sont alors décrites par des degrés d'appartenance qui indiquent à quel point leurs caractéristiques appartiennent à la modalité considérée. Le tableau 1 présente l'exemple d'une telle base, contenant 8 objets décrits par 3 attributs. Ainsi, la vitesse du premier objet appartient avec un degré 0.5 à la modalité rapide, avec un degré 0.3 à normale et un degré 0.2 à faible.

2.1 Définitions de motifs et règles graduels

On définit alors les notions d'item graduel, motif (ou itemset) graduel, et règle graduelle de la manière suivante [BCS+07, DLT09] :

Définition 1 (Item graduel) *Soit X un attribut défini sur un univers U, et V un sens de variation qui peut être croissant ($plus$) ou décroissant ($moins$). On appelle* item graduel *le couple (X, V).*

Ainsi, *(vitesse, plus)* et *(vitesse, moins)* sont des items graduels. Dans le cas de schéma de données flou, on définit un item graduel flou de la façon suivante [BCS+07, DLT09] :

Définition 2 (Item graduel flou) *Soit X un attribut défini sur un univers U correspondant à une variable linguistique et A un sous-ensemble flou de U décrivant l'une des valeurs linguistiques. On appelle* item graduel flou *le couple (X, A).*

Il faut noter que le sens de variation n'est alors plus décrit et qu'il est croissant par défaut : ainsi l'item graduel flou *(vitesse, rapide)* peut

Identifiant	Vitesse			Distance Mur		Freinage		
	Faible	Normale	Rapide	Proche	Très Proche	Léger	Moyen	Fort
o_1	0.2	0.3	0.5	0.4	0.6	0.6	0.4	0.2
o_2	0.2	0.2	0.6	0.5	0.5	0.2	0.7	0.1
o_3	0	0.1	0.9	0.7	0.3	0	0.6	0.4
o_4	0	0.2	0.8	0.8	0.2	0.1	0.3	0.6
o_5	0.1	0.7	0.3	0.3	0.7	0.3	0.3	0.4
o_6	0.2	0.3	0.5	0.9	0.1	0.5	0.3	0.2
o_7	0	0.6	0.4	0.8	0.2	0.1	0.1	0.8
o_8	0.1	0.2	0. 7	0.9	0.1	0	0.3	0.7

Tableau 1 – Exemple de base de données floues

être lu comme *plus la vitesse est rapide* et interprété comme imposant la contrainte que les degrés d'appartenance à la modalité *rapide* soient croissants.

La combinaison de plusieurs items graduels se fait alors au sein d'un motif graduel, également nommé itemset graduel.

Définition 3 (Motif graduel) *Un* motif graduel *est défini comme un ensemble* $\{i_1, \ldots, i_n\}$ *d'items graduels. Si les items sont flous, le motif est dit* motif graduel flou.

On notera que l'ensemble décrit par un motif est sémantiquement associé à une conjonction. Ainsi, {*(vitesse, plus),(distanceMur, moins)*} est un motif graduel, et {*(vitesse, rapide),(distanceMur, proche)*} est un motif graduel flou, le premier motif signifiant qu'on considère *(vitesse, plus)* **et** *(distanceMur, moins)*.

On définit ensuite les règles graduelles et leurs variantes floues de la façon suivante :

Définition 4 (Règle graduelle) *Soient* M_1 *et* M_2 *deux motifs graduels, une* règle graduelle *est de la forme* $M_1 \rightarrow M_2$. M_1 *est l'*antécédent *et* M_2 *le* conséquent.

Une *règle graduelle floue* est une règle graduelle composée de motifs graduels flous. On notera ci-dessous *rgf* pour règle graduelle floue. Elle peut par exemple prendre la forme *plus le mur est proche, plus le freinage est fort.*

2.2 Définition de support et confiance

La pertinence d'une règle graduelle est alors évaluée par son support et sa confiance. Plusieurs approches ont été proposées pour définir ces mesures, nous rappelons et utilisons ici les définitions proposées dans [DLT08, DLT09] :

Définition 5 (Support d'un motif graduel) *Soit* DB *une instance de base de données composée d'un ensemble de lignes (*m*-uplets) et* $M = (X_1, V_1) \ldots (X_n, V_n)$ *un motif graduel. Le* support du motif M *est le nombre maximal de lignes* $\{l_1, \ldots, l_p\}$ *de* DB *pour lesquelles il existe une permutation* π *avec* $\{l_{\pi_1}, \ldots, l_{\pi_p}\}$ *telle que* $\forall j \in [1, p-1]$, $\forall k \in [1, n]$, *on a* $l_{\pi_j}[X_k] \leq l_{\pi_{j+1}}[X_k]$ *si* V_k *= plus et* $l_{\pi_j}[X_k] \geq l_{\pi_{j+1}}[X_k]$ *sinon.*

Dans le cas d'un motif graduel flou, on considère la définition précédente avec un sens de variation V_i toujours croissant et X_i correspondant à la valeur linguistique considérée. La contrainte de croissance est alors imposée aux degrés d'appartenance à cette modalité.

Le support d'une règle est ensuite défini comme le support de l'union des motifs qu'elle fait intervenir :

Définition 6 (Support d'une règle graduelle) *Soient* M_1 *et* M_2 *deux motifs graduels, le* support de la règle graduelle $R = M_1 \rightarrow M_2$ *est calculé comme* $supp(R) = supp(M_1 \bigcup M_2)$.

Le support compte le nombre maximal de lignes de la base de données qui peuvent être ordonnées de manière à ce que chaque ligne respecte les relations sur tous les attributs X_i et Y par rapport aux lignes voisines. La confiance d'une règle est ensuite définie comme la proportion de lignes respectant la gradualité imposée par le conséquent de la règle parmi celles qui respectent la gradualité imposée par l'antécédent :

Définition 7 (Confiance d'une règle graduelle) *Soient* M_1 *et* M_2 *deux motifs graduels, la* confiance de la règle graduelle $R = M_1 \rightarrow M_2$ *est définie comme* $conf(R) = \frac{supp(M_1 \bigcup M_2)}{supp(M_1)}$.

La confiance donne une indication sur la force du lien de causalité tandis que le support donne une indication sur la proportion de la base de données concernée par les motifs mis en jeu. Le but des méthodes d'extraction est de déduire automatiquement des bases de données les règles graduelles floues apparaissant au-delà d'un support et d'une confiance définis par l'utilisateur.

2.3 Algorithmes d'extraction

Les méthodes existantes d'extraction de règles graduelles diffèrent dans la manière de retrouver les motifs et les règles dont le support (éventuellement la confiance) est supérieur à un seuil fixé par l'utilisateur (on les dit alors *fréquents*), certaines utilisant des définitions de support légèrement différentes de la définition 5.

Dans [Hül02], l'auteur propose une approche basée sur des méthodes de régression pour trouver les règles graduelles.

Dans [BCS+07], les auteurs considèrent une approche par niveau, mettant en relation deux, puis trois, puis quatre, etc. attributs. Le support d'un motif graduel est calculé en comparant les lignes de la base de données deux à deux afin de calculer le nombre de couples de lignes respectant le motif considéré. Cette méthode est très coûteuse en terme de temps de calcul, les expérimentations montrées par les auteurs se limitent donc à quelques attributs seulement.

Dans [LLR09] les auteurs considèrent la même définition de support en fonction des couples de données respectant le motif, qu'ils interprètent en termes de corrélation d'ordre, mesurée par le taux de Kendall. Ils proposent une méthode efficace de calcul basée sur une représentation binaire des propriétés des couples d'objets.

Dans [DLT08], les auteurs proposent une heuristique permettant de retrouver le support d'un motif graduel en utilisant une approche par niveau. Cette heuristique consiste à éliminer de la base de données, à chaque niveau, les lignes dont l'ensemble de conflit est maximal, définies comme les lignes qui empêchent le plus de lignes d'être ordonnées. Il s'agit d'une heuristique dans le sens où le choix d'une autre ligne à un niveau donné aurait pu mener à de meilleurs résultats au niveau suivant.

Dans [DLT09], les auteurs proposent aussi une approche par niveau, qui permet le calcul du support et la génération des motifs graduels de manière très performante : les ordonnancements de lignes sont représentés sous la forme d'un graphe dont les nœuds sont les lignes de la base et les arêtes sont les liens de précédence par rapport à tous les attributs impliqués dans le motif considéré. Retrouver le chemin maximal dans ce graphe permet alors de retrouver le support du motif considéré. Cette représentation est stockée en mémoire sous forme binaire, ce qui permet de calculer rapidement des jointures entre motifs graduels et de passer efficacement d'un niveau au suivant.

3 Règles graduelles floues renforcées

Nous définissons dans cette section la notion de règle graduelle floue renforcée, du type *plus* *la distance au mur est proche, plus le freinage est fort*, d'autant plus que *la vitesse est élevée*. Nous discutons la définition du renforcement exprimé par la clause *d'autant plus que* puis proposons des mesures de support et confiance qui permettent d'évaluer la qualité des règles renforcées. Enfin nous présentons un algorithme permettant d'extraire automatiquement ces règles à partir de bases de données.

3.1 Définition et sémantique

Formellement, on peut donner la définition suivante des règles graduelles floues renforcées et de leurs composantes :

Définition 8 (Règle graduelle floue renforcée) *Une* règle graduelle floue renforcée *est un triplet de motifs graduels flous (M_1, M_2, M_3) noté* $M_1 \rightarrow M_2; M_3$. M_1 *est l'*antécédent*,* M_2 *le* conséquent*, et* M_3 *le* renforcement*, exprimé par la clause* d'autant plus que.

Ainsi, avec $M_1 =$ *(vitesse, élevée)*, $M_2 =$ *(danger, grand)* et $M_3 =$ *(virages, serrés)*, la règle graduelle floue renforcée $M_1 \rightarrow M_2; M_3$ représente l'exemple : *plus la vitesse est élevée, plus le danger est grand, d'autant plus que les virages sont serrés*.

Il est important de souligner la différence entre ces règles renforcées et des règles conjonctives de la forme $(M_1 \wedge M_3) \rightarrow M_2$. Dans le cas de l'exemple précédent, une telle règle conjonctive s'écrit *plus la vitesse est élevée et plus les virages sont serrés, plus le danger est grand*. M_3 est alors identifié comme l'une des causes de M_2, alors que ce n'est pas le cas pour la règle renforcée. De plus, la règle conjonctive exprime une forte contrainte puisqu'elle impose l'identité des ordonnancements des données selon les trois motifs M_1, M_2 et M_3. Dans le cas du renforcement, la contrainte d'ordonnancement concerne seulement deux motifs, M_1 et M_2, elle restreint donc moins le nombre d'objets concernés par la règle. L'apport du troisième motif, de renforcement, se fait selon une autre sémantique, qui ne s'exprime pas strictement en termes d'ordre.

Différentes possibilités pour l'interprétation de la relation de renforcement peuvent être envisagées. Ainsi, on peut traduire l'influence de M_3 comme une contrainte sur la rapidité de variations de M_2 lorsque l'identité des ordonnancements selon les motifs M_1 et M_2 est acquise : la relation "d'autant plus que" est comprise comme une accélération des variations des attributs impliqués dans le motif M_2 en fonction des valeurs prises par les attributs qui interviennent dans M_3. Sur l'exemple précédent, cela signifie que d'une part une différence de vitesse produit une différence de dangerosité et que, d'autre part, cette différence est plus importante lorsqu'une différence des caractéristiques des virages est observée. Cette interprétation de renforcement peut être exprimée comme la combinaison de plusieurs règles graduelles floues : la règle $M_1 \rightarrow M_2$, et une règle du type $M_3 \rightarrow \Delta M_2$, en notant ΔM_2 les variations de M_2.

Une autre interprétation, que nous adoptons dans cet article, évalue le renforcement induit par le motif M_3 sur la règle $M_1 \rightarrow M_2$ selon le nombre de lignes de la base de données qui vérifient la règle en examinant à quel point M_3 est possédé par les objets qui vérifient la règle. On considère alors dans quelle mesure $M_1 \rightarrow M_2$ est vraie sur toute la base de données, et on compare ce résultat avec le point auquel cette même règle est vraie quand on se restreint aux cas où M_3 est vrai. La clause de renforcement M_3 pouvant être floue, nous utilisons le degré de vérité de M_3 et les cardinalités floues pour nous retreindre à cette sous-base.

3.2 Support et confiance

Nous décrivons ici les définitions des mesures de qualité, support et confiance, proposées pour évaluer les règles graduelles floues renforcées selon la sémantique précisée ci-dessus. Une règle de la forme $M_1 \rightarrow M_2; M_3$ comprenant deux composantes, la règle $M_1 \rightarrow M_2$, et son renforcement, elle doit être évaluée selon les deux aspects. Aussi nous proposons de caractériser une règle renforcée à la fois par le support et la confiance de la règle $M_1 \rightarrow M_2$, et par des notions de support et de confiance dans le renforcement, qui mesurent la qualité de ce dernier.

Comme dans le cas classique, nous proposons de définir le support de la règle renforcée par une évaluation de sa fréquence. Pour cela, nous proposons d'utiliser une procédure de sigma-comptage pour définir le degré avec lequel M_3 est présent parmi les objets qui vérifient la règle $M_1 \rightarrow M_2$: plus il est élevé, plus la règle renforcée est valide. Plus formellement, la définition est la suivante :

Définition 9 (Support d'une rgf renforcée)
Soit $R = M_1 \rightarrow M_2; M_3$ une règle graduelle floue renforcée et $\mathcal{L} = \{L_1, \ldots, L_k\}$ l'ensemble des ensembles de lignes pouvant être ordonnés pour respecter l'ordre sur M_1 et M_2.
Soit $\ell[M_3]$ le degré d'appartenance d'une ligne ℓ à M_3.
On note $s(L_i) = \sum_{\ell \in L_i} \ell[M_3]$ le sigma-comptage correspondant à l'ensemble de lignes L_i.
Le support de la règle graduelle floue renforcée *R est alors calculé comme*

$$r\_supp(R) = \max_{i=1,\ldots,k} s(L_i)$$

À titre d'exemple, considérons les données présentées dans le tableau 1 et la règle R définie par *plus le mur est proche alors plus le freinage est fort, d'autant plus que la vitesse est rapide*. Le support renforcé de R est $r\_supp(R) = 3$. En effet, on a $\mathcal{L} = \{\{1,3,4,7\}, \{2,3,4,7\}, \{1,3,4,8\}, \{2,3,4,8\}\}$, $s(\{1,3,4,7\}) = 0.5+0.9+0.8+0.4 = 2.6$, $s(\{2,3,4,7\}) = 0.6+0.9+0.8+0.4 = 2.7$, $s(\{1,3,4,8\}) = 0.5+0.9+0.8+0.7 = 2.9$, et $s(\{2,3,4,8\}) = 0.6+0.9+0.8+0.7 = 3$.

Comme indiqué dans la définition 3, un motif est associé à la conjonction des items qu'il contient. Aussi, si M_3 est un motif complexe, comportant plusieurs items, le degré d'appartenance à M_3, $\ell[M_3]$, est calculé par l'agrégation par une t-norme des degrés d'appartenance à chacun des items.

Il faut noter que la définition précédente est asymétrique, du fait du rôle spécifique de M_3. Par ailleurs, comme discuté dans la section précédente, elle ne prend pas en compte une co-variation de la clause de renforcement avec l'antécédent et le conséquent de la règle : elle utilise un sigma-comptage sur la présence de M_3. Cette approche est en particulier intéressante pour le traitement du renforcement par des attributs non flous comme par exemple *plus l'achat de viande est élevé, plus l'achat de légumes est moyen, d'autant plus que la ville est Paris*, qui peut être aisément traité par ce biais, le degré d'appartenance étant alors égal à 1 pour toutes les lignes de la base de données possédant l'attribut.

On peut remplacer dans la définition précédente le sigma-comptage par un sigma-comptage seuillé, en utilisant un seuil défini par l'utilisateur : cela permet d'éviter de prendre en compte des lignes qui ne seraient pas assez représentatives de la clause M_3.

Le support renforcé r_supp possède une propriété d'anti-monotonie relativement à la longueur du motif M_3 : si on considère deux règles, $R_1 = M_1 \rightarrow M_2; M_3$ et $R_2 = M_1 \rightarrow M_2; M_4$ avec $M_3 \subset M_4$, on a $r\_supp(R_1) \geq r\_supp(R_2)$. En effet, les deux règles R_1 et R_2, ayant la même forme non renforcée, ont le même ensemble $\mathcal{L} = \{L_1, ..., L_k\}$ de lignes pouvant être ordonnées pour respecter l'ordre sur M_1 et M_2. De plus, $\forall i \in [1,k]$, $\forall \ell \in L_i$, $\ell[M_3] \geq \ell[M_4]$: comme un motif est associé à la conjonction des items qu'il contient, en notant M l'ensemble d'items tel que $M_4 = M_3 \cup M$ et $\top$ une t-norme, on a $\ell[M_4] = \top(\ell[M_3], \ell[M]) \leq \ell[M_3]$. On a donc $\forall i$, $s_{R_1}(L_i) \geq s_{R_2}(L_i)$, ce qui conduit à la relation d'anti-monotonie.

La notion de confiance de la règle renforcée évalue ensuite la pertinence du renforcement par rapport à la règle non renforcée $M_1 \rightarrow M_2$: suivant le schéma classique de définition de confiance, qui rapporte le support de la règle au support de l'antécédent, nous proposons de comparer le support de la règle renforcée, tel que défini ci-dessus, au support de la règle non renforcée.

Cette définition est équivalente au calcul du quotient entre le cardinal flou, calculé selon le motif M_3, de l'ensemble des lignes compatibles avec la règle $M_1 \rightarrow M_2$, par le nombre de ces lignes. Plusieurs ensembles de telles lignes pouvant exister, nous considérons pour la normalisation l'ensemble qui maximise le cardinal flou. Plus formellement, en utilisant les mêmes notations que dans la définition 9, et en introduisant la notation $i* = \arg\max_{i=1,\dots,k} s(L_i)$, nous posons :

Définition 10 (Confiance d'une rgf renforcée) *La* confiance de la règle graduelle floue renforcée $R = M_1 \rightarrow M_2; M_3$ *est calculée comme*

$$\begin{aligned} r\_conf(R) &= \frac{r\_supp(M_1 \rightarrow M_2; M_3)}{supp(M_1 \rightarrow M_2)} \\ &= \frac{s(L_{i*})}{|L_{i*}|} \end{aligned}$$

En reprenant l'exemple précédent, la confiance de la règle *plus le mur est proche alors plus le freinage est fort, d'autant plus que la vitesse est rapide* est alors égale à 3/4 = 0.75.

Il faut noter que, comme le support renforcé, la confiance renforcée possède une propriété d'anti-monotonie relativement à la longueur du motif M_3 : en notant $R_1 = M_1 \rightarrow M_2; M_3$ et $R_2 = M_1 \rightarrow M_2; M_4$ avec $M_3 \subset M_4$ $r\_conf(R_1) \geq r\_conf(R_2)$. En effet, dans la définition ci-dessus, R_1 et R_2 ayant la même forme non renforcée, elles ont le même support, donc le même dénominateur. Or le support renforcé au numérateur vérifie la propriété d'anti-monotonie, comme indiqué précédemment.

3.3 Propriétés et algorithme

Le problème est alors d'extraire toutes les règles graduelles floues renforcées dont le support, la confiance, le support renforcé et la confiance renforcée sont supérieurs à des seuils fixés par l'utilisateur. Notons que les règles recherchées sont les règles dites maximales, c'est-à-dire celles qui contiennent le plus grand nombre possible d'items graduels tout en respectant les seuils de support, confiance et renforcement. L'exigence de maximalité s'applique aux trois motifs constituant respectivement la prémisse, la conclusion et le renforcement.

Une règle graduelle floue renforcée $M_1 \rightarrow M_2; M_3$ comporte deux composantes, la règle graduelle floue $M_1 \rightarrow M_2$ et le renforcement M_3, la première devant vérifier les contraintes de confiance et support indépendamment de la seconde. Aussi l'algorithme est constitué de deux étapes : la première consiste à identifier les règles graduelles floues vérifiant les contraintes de confiance et support, la seconde vise à établir les renforcements qui peuvent leur être adjoints.

La première étape s'effectue grâce à l'algorithme GRITE [DLT09] qui extrait les motifs graduels flous fréquents et leurs supports et qui permet de plus d'identifier les ensembles de lignes de la base de données qui peuvent être ordonnées pour respecter les contraintes d'ordre exprimées par ces motifs. L'algorithme utilise une représentation efficace des relations de précédence entre les lignes de données sous la forme de matrice d'adjacence binaire pour fournir la totalité de ces informations avec une complexité de temps de calcul très faible.

On peut noter que les définitions de support et de confiance renforcés (définitions 9 et 10 respectivement) dépendent en fait des motifs fréquents $M_1 \cup M_2$ et non des règles $M_1 \rightarrow M_2$. Elles ne nécessitent donc pas l'identification des relations de causalité parmi les items constituant les motifs considérés.

La seconde étape qui vise à établir les renforcements pour ces motifs suit le principe de l'algorithme APRIORI [AS94] transposé pour extraire les motifs de renforcement et non les motifs fréquents : les renforcements sont établis selon une approche par niveau, construisant progressivement les motifs de taille croissante, exploitant les propriétés d'anti-monotonie du support renforcé et de confiance renforcée.

Pour cela, étant donné une règle graduelle floue R à renforcer (ou un motif graduel flou fréquent), les motifs de renforcement candidats de taille $k + 1$ sont générés à partir des motifs candidats de taille k selon la procédure classique APRIORIgen. Ils sont ensuite évalués en utilisant les mesures de support renforcé et de confiance renforcée. Les candidats pour lesquels l'un ou l'autre des critères a une valeur inférieure aux seuils fixés par l'utilisateur sont éliminés. Ces deux étapes sont itérées jusqu'à épuisement des candidats.

Lors de la phase d'initialisation, les motifs de renforcement candidats contenant un seul item sont définis comme les modalités des attributs dont aucune modalité n'apparaît dans la règle (ou le motif) que l'on cherche à renforcer. En effet, un même attribut ne peut apparaître sous deux modalités différentes dans une même règle.

4 Conclusion

Dans cet article, nous proposons une nouvelle approche permettant de définir les règles graduelles floues renforcées en regard des dernières avancées de la littérature concernant les règles graduelles et la fouille de données. Les algorithmes associés sont décrits, ils s'inscrivent dans le contexte des méthodes d'extraction par niveau.

Les perspectives de ce travail concernent notamment l'application de cette approche à des données réelles ainsi que la comparaison des résultats avec les méthodes existantes basées sur les approches de fouille de données.

Références

[AS94] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proc. of VLDB'94*, pages 487–499, 1994.

[BCS+07] F. Berzal, J.-C. Cubero, D. Sanchez, M.-A. Vila, and J. M. Serrano. An alternative approach to discover gradual dependencies. *IJUFKS*, 15(5) :559–570, 2007.

[Del99] J. Delechamp. *Un modèle flou de variations graduelles autour de prototypes*. PhD thesis, Université Paris 6, 1999.

[DLT08] L. Di Jorio, A. Laurent, and M. Teisseire. Fast extraction of gradual association rules : A heuristic based method. In *Proc. of CSTST'08*, 2008.

[DLT09] L. Di Jorio, A. Laurent, and M. Teisseire. Mining frequent gradual itemsets from large databases. In *Proc. of IDA'09*, 2009.

[DP92] D. Dubois and H. Prade. Gradual inference rules in approximate reasoning. *Information Sciences*, 61(1-2) :103–122, 1992.

[GDP04] S. Galichet, D. Dubois, and H. Prade. Imprecise specification of ill-known functions using gradual rules. *IJAR*, 35 :205–222, 2004.

[Hül02] E. Hüllermeier. Association rules for expressing gradual dependencies. In *Proc. of PKDD'02*, pages 200–211, 2002.

[LLR09] A. Laurent, M.-J. Lesot, and M. Rifqi. Graank : Exploiting rank correlations for extracting gradual dependencies. In *Proc. of FQAS'09*, 2009.