Préférences et quantités dans le cadre de l'interrogation flexible : sur la prise en compte d'expressions quantifiées

Daniel Rocacher[1] Ludovic Liétard[2]

[1] IRISA/ENSSAT
[2] IRISA/IUT Lannion

[1] BP 447, 22305 Lannion Cédex, France, rocacher@enssat.fr
[2] BP 150, 22302 Lannion cédex, France, ludovic.lietard@univ-rennes1.fr

Résumé :
L'interrogation flexible permet de prendre en compte des préférences de l'utilisateur dans les requêtes. Ces préférences sont exprimées au moyen de critères de sélection définis par des ensembles flous. Le traitement conjoint des notions de préférence et de cardinalité nous a conduits à définir le concept d'entier graduel ($\mathbb{N}_f$). Ce cadre a ensuite été étendu (en $\mathbb{Z}_f$ et $\mathbb{Q}_f$) afin de pouvoir traiter des requêtes comportant des différences et des divisions. Dans ce papier, nous étudions comment définir des relations d'ordre entre ces nombres graduels. Puis nous montrons que ces concepts appliqués au traitement d'expressions quantifiées du type "Q X sont A" ou "Q B X sont A" fournissent un nouveau cadre pour leurs interprétations.

Mots-clés :
Nombres graduels, relation d'ordre, conjonction, disjonction, interrogation flexible.

Abstract:
Based on fuzzy set theory, flexible querying enables users to express preferences inside requirements. Quantifications and preferences on data have led us to define the notion of fuzzy integers ($\mathbb{N}_f$). This framework has been extended (to $\mathbb{Z}_f$ and $\mathbb{Q}_f$) in order to dealing with queries based on difference or division operations. In this paper we study how to define fuzzy order relations between such fuzzy numbers. This approach applies to quantified statements of type "Q X are A" and "Q B X are A" and provides a new theoretical background for their interpretations.

Keywords:
Gradual numbers, order relation, conjonction, disjonction, flexible querying.

1 Introduction

Un des objectifs de la recherche dans le domaine des bases de données est d'améliorer la capacité d'expression des langages de requêtes. Un des moyens étudiés est la prise en compte de préférences afin de faciliter l'accès à des informations pertinentes.

Différentes propositions ont été faites pour introduire des préférences dans des requêtes. On peut distinguer deux approches générales selon que les valeurs de préférences associées aux attributs sont commensurables ou pas. Dans le premier cas, les valeurs de préférences peuvent être agrégées pour délivrer une valeur globale et définir un ordre total sur les réponses. Dans le second cas, lorsqu'il n'y a pas commensurabilité, seul un ordre partiel des réponses, basé sur l'ordre de Pareto, est possible et des classes incomparables de réponses sont construites. Cette approche est détaillée dans [7] et illustrée par l'opérateur Skyline [3] ou dans PreferenceSQL [14].

Les préférences des utilisateurs peuvent aussi être exprimées par des critères de sélection fondés sur des ensembles flous. Les prédicats ne sont alors plus en « tout ou rien » mais peuvent être plus ou moins satisfaits. Par exemple, dans la requête « trouver les employés jeunes et bien-payés de l'entreprise *X* », les critères *jeune* et *bien-payé* peuvent être plus ou moins satisfaits et leur définition doit tenir compte de préférences sur les valeurs des domaines indiquant lesquelles correspondent le mieux aux concepts décrits. Les critères *jeunes* et *bien-payé* sont définis par des ensembles flous permettant de restituer pour un âge et un salaire donnés des satisfactions définies sur une *même* échelle d'interprétation, généralement l'intervalle [0, 1]. Ces degrés de satisfaction sont donc *commensurables* et, en conséquence, composables.

Ainsi, la théorie des ensembles flous est un cadre général permettant d'exprimer des requêtes flexibles. Il a été montré [23] qu'elle généralise d'autres propositions essentiellement basée sur des distances [12][17][22]. Un ensemble flou *E* est défini par une fonction caractéristique μ à valeur dans l'intervalle [0, 1] de $\mathbb{R}$, telle que $\mu_E(x)$ exprime dans quelle mesure l'élément x appartient à l'ensemble flou *E*. Les prédicats dits graduels (*i.e.* dont le résultat est un degré de satisfaction), comme *jeune* et *bien-payé*, sont décrits au moyen d'ensembles flous. Ces critères peuvent être combinés grâce à des opérateurs de conjonction ou de disjonction (éventuellement pondérés pour exprimer des importances relatives entre critères) ou de moyennes exprimant des effets de compensation entre critères. Les résultats d'une requête flexible sont alors qualifiés en fonction de leur adéquation aux critères de sélection et peuvent être ordonnés. Divers travaux portant sur l'expression et l'interprétation de requêtes flexibles dans le contexte du modèle relationnel [6][13] et du modèle objet [8] ont vu le jour. Cette étude se place dans ce contexte de l'interrogation flexible de bases de données usuelles (*i. e.* : ne contenant pas de données mal connues) et s'intéresse plus particulièrement au traitement conjoint de préférences et de quantités sur les données manipulées.

Concernant le traitement de quantités associées à des données, le concept de multi-ensemble, c'est-à-dire de collection autorisant des occurrences multiples de ses éléments, est un support intéressant. Il trouve de nombreuses applications, en algorithmique notamment [16], mais aussi dans les bases de données [1][18]. Dans ce contexte, l'utilisation des multi-ensembles est principalement motivée par leur capacité à gérer des quantités. On note qu'ils facilitent également le traitement de certaines opérations en évitant la suppression des doubles. Les propriétés algébriques des multi-ensembles ont été largement étudiées, voir [2] pour une présentation complète.

Nous avons montré [25][30] que l'utilisation des multi-ensembles pour gérer des quantités et des ensembles flous pour gérer des préférences, conduit à la définition d'une généralisation de ces structures appelée multi-ensemble flou [21][34]. Un multi-ensemble flou est un multi-ensemble dont chaque occurrence d'un élément est associée à un degré d'appartenance. Par exemple, un multi-ensemble flou peut être délivré par la requête : *trouver le salaire des employés jeunes*. Comme plusieurs employés peuvent avoir le même salaire, le résultat peut contenir des doubles. De plus, chaque occurrence d'un salaire coïncide avec un employé plus ou moins jeune et est donc associée à un degré de satisfaction. Le multi-ensemble flou résultat correspond alors *la distribution des salaires des employés jeunes*.

Dans [25] nous avons proposé une approche qui caractérise les multi-ensembles flous et permet de traiter de manière uniforme les ensembles, ensembles flous, multi-ensembles et multi-ensembles flous. Cette construction s'appuie sur le concept d'entier naturel graduel ($\mathbb{N}_f$) qui correspond à la cardinalité d'un ensemble flou. Par la suite [28], nous avons défini un cadre plus général, basé sur l'ensemble des entiers relatifs graduels ($\mathbb{Z}_f$), dans lequel il est possible de définir des différences exactes. L'intérêt de cette démarche est d'offrir une base algébrique permettant la composition de calculs. Enfin, $\mathbb{Z}_f$ a été prolongé en $\mathbb{Q}_f$ [26][27], l'ensemble des nombres rationnels graduels, afin de construire un système d'opérations multiplicatives fermé et de réaliser des divisions exactes. Ces contextes permettent de traiter des requêtes flexibles complexes basées sur des calculs entre quantités graduelles ou nécessitant la manipulation de multi-ensembles flous [24]. Des requêtes de ce type sont par exemple :

- calculer la *moyenne* des salaires des employés *jeunes* ;
- trouver les entreprises où le *nombre des employés proches de la retraite* est plus grand que *celui des employés jeunes ;*
- trouver les entreprises dont *la plupart* des employés *jeunes* sont *bien-payés.*

Dans cet article, nous nous intéressons à l'utilisation de ces notions pour traiter des requêtes portant sur des quantités graduelles. Ainsi, notre objectif est d'étudier des conditions flexibles telles que : *environ deux employés jeunes sont bien payés* ou *la plupart des employés jeunes sont bien payés* qui comportent respectivement un quantificateur absolu (*environ deux*) et un quantificateur relatif (*la plupart*). Pour évaluer ces conditions il est nécessaire de définir des relations d'ordre sur des quantités graduelles.

Par la suite, en section 2, les constructions de $\mathbb{N}_f$, $\mathbb{Z}_f$ et $\mathbb{Q}_f$ sont brièvement rappelées. Puis, en section 3, nous étudions comment définir des relations d'ordre entre quantités graduelles au moyen d'une intégrale pondérée. Nous montrons que cette notion de relation d'ordre généralise le concept d'implication floue définie dans le cadre de la logique multivaluée. Le cadre ainsi obtenu nous permet, en section 4, de proposer un mécanisme de base pour traiter des requêtes flexibles comportant des expressions quantifiées.

2 Quantités graduelles

Dans cette section nous décrivons quelques éléments de base relatifs à la modélisation de quantités graduelles. Ceux-ci sont nécessaires à la compréhension des sections suivantes qui traitent de la comparaison de quantités graduelles et à leurs utilisations dans des expressions quantifiées.

2.1 Entiers naturels graduels

Un ensemble flou, défini sur un domaine X, peut se définir par une fonction caractéristique, notée μ_E, telle que :

$$\mu_E : X \rightarrow [0, 1]$$
$$x \rightarrow \mu_E(x).$$

La valeur $\mu_E(x)$ exprime dans quelle mesure l'élément x de X appartient à l'ensemble flou E. Quand $\mu_E(x)$ est nul, x n'appartient pas du tout à E et quand il vaut 1, x est complètement dans E. Plus $\mu_E(x)$ est proche de 1 (resp. 0), plus (resp. moins) x appartient à E. Dans le cas d'un ensemble flou fini E, on note généralement :

$$E = \{\mu_E(x_1)/x_1, \ldots, \mu_E(x_n)/x_n\} = \sum_{i=1}^{n} \mu_E(x_i)/x_i \,.$$

Un ensemble flou peut également être décrit comme une collection d'ensembles ordinaires emboîtés grâce à la notion de *coupes de niveau* ou *α-coupes*. La coupe de niveau α de l'ensemble flou E, notée E_α, est l'ensemble usuel composé des éléments dont le degré d'appartenance à E est au moins égal à α, d'où :

$$E_\alpha = \{x \,/\, x \in X \text{ et } \mu_E(x) \geq \alpha\}.$$

La cardinalité floue $|E|$ d'un ensemble flou E est définie par un ensemble flou d'entiers caractérisé par :

$$\forall\, n \in \mathbb{N},\ \mu_{|E|}(n) = sup\{\alpha \;/\; |E_\alpha| \geq n\}.$$

Cette définition est aussi appelée FGCount(E) par Zadeh [36]. Le degré α, associé à un entier n de $|E|$, évalue dans quelle mesure E contient au moins n éléments.

Exemple 2.1.

Soit l'ensemble flou $E = \{1/x_1, 1/x_2, 0.5/x_3, 0.2/x_4\}$, la cardinalité de E est représentée par :

$|E| = \{1/0, 1/1, 1/2, 0.5/3, 0.2/4\}.$

Le degré 0.5 de $|E|$ exprime dans quelle mesure l'ensemble flou E contient au moins 3 éléments.
♦

La cardinalité d'un ensemble usuel fini peut être vue comme un entier naturel n. La cardinalité floue d'un ensemble flou fini peut être vue comme un entier que nous qualifions de graduel pour le distinguer de la notion de nombre flou au sens habituel du terme [10]. En effet, il est important de noter que nous adoptons ici un point de vue nouveau interprétant une cardinalité floue comme "un tout" décrivant complètement, et *exactement*, la cardinalité d'un ensemble flou. Ce nombre est traité comme un ensemble flou *conjonctif* d'entiers décrivant la cardinalité d'un ensemble flou selon ces différents niveaux d'exigence. Il est parfaitement connu et ne recouvre aucune forme d'incertitude. De ce point de vue, il se différencie de la définition habituelle de nombre flou qui est interprété comme une valeur mal connue modélisée par une distribution de possibilité (correspondant à une collection *disjonctive* de valeurs possibles) [9]. Comme nous le verrons par la suite, ces spécificités ont des conséquences importantes sur les propriétés algébriques des opérations associées à ces nombres. Nous appelons $\mathbb{N}_f$ l'ensemble des entiers naturels graduels.

L'α-coupe d'un entier graduel x est un ensemble d'entiers formant une suite croissante $\{0, 1, \ldots, x_\alpha\}$ qui peut être représentée par sa plus grande valeur x_α. Par la suite, on appelle coupe de niveau α d'un entier graduel x le plus grand entier de l'α-coupe de l'ensemble flou d'entiers défini par x. L'α-coupe d'un entier graduel x est donc interprétée comme un entier positif que l'on note x_α.

Nous avons montré que cette notion d'entier graduel permet de définir le concept de multi-ensemble flou, c'est-à-dire de multi-ensemble dont les occurrences des différents éléments sont associées à un degré d'appartenance. En effet, un élément x, dans un multi-ensemble flou A, est caractérisé par un entier graduel, noté $\Omega_A(x)$, représentant la cardinalité de l'ensemble flou de des différents occurrences de x dans A. Les opérations de composition des multi-ensembles flous (intersection, union, union additive, produit cartésien, …) peuvent ensuite être définies à partir d'opérations de base sur les entiers graduels (min, max, $\oplus$, $\otimes$, …). La définition d'une opération binaire * entre deux entiers graduels Q et Q' s'appuie sur le principe d'extension étendu [33] et est définie par :

$$\mu_{Q*Q'}(z) = \sup_{(x,y)/x*y\geq z} \min(\mu_Q(x), \mu_{Q'}(y))$$

Les opérations * sur les entiers graduels respectent la propriété caractéristique suivante :

$$(x * y)_\alpha = x_\alpha * y_\alpha.$$

Exemple 2.2.

Soit le multi-ensemble flou suivant : $A = \{<1, 0.1, 0.1>/a, <0.5>/b\}$ qui signifie que A contient trois occurrences de l'élément a, chacune étant affectée d'un degré d'appartenance, respectivement 1, 0.1 et 0.1, et une occurrence de b au degré 0.5.

Les nombres graduels d'occurrences des éléments sont :

$\Omega_A(a) = \{1/0, 1/1, 0.1/2, 0.1/3\}$;

$\Omega_A(b) = \{1/0, 0.5/1\}$.

A peut également se noter : $A = \{\{1/0, 1/1, 0.1/2, 0.1/3\}*a, \{1/0, 0.5/1\}*b\}$.

Si B est le multi-ensemble flou $\{\{1/0, 1/1, 0.5/2\}*a, \{1/0, 0.5/1\}*b\}$. L'union additive $A + B$, obtenue en regroupant les éléments de A et de B, est définie par :

$\Omega_{A+B}(a) = \Omega_A(a) \oplus \Omega_B(a) = \{1/0, 1/1, 0.1/2, 0.1/3\} \oplus \{1/0, 1/1, 0.5/2\} = \{1/0, 1/1, 1/2, 0.5/3, 0.1/4, 0.1/5\}$

$\Omega_{A+B}(\mathrm{b}) = \Omega_A(b) \oplus \Omega_B(b) = \{1/0, 0.5/1\} \oplus \{1/0, 0.5/1\} = \{1/0, 0.5/1, 0.5/2\}$.

♦

2.2 Entiers relatifs graduels

Sur $\mathbb{N}_f$, la différence de deux multi-ensembles flous peut ne pas toujours être définie. C'est pourquoi $\mathbb{N}_f$ a été étendu en $\mathbb{Z}_f$, l'ensemble des entiers relatifs graduels. Nous avons montré que ce support permet une généralisation où la différence peut toujours être représentée. Nous rappelons maintenant brièvement comment nous avons construit l'ensemble des entiers relatifs graduels.

Soit a et b deux entiers naturels graduels et $\mathcal{R}$ la relation d'équivalence telle que :

$\forall (a, b) \in \mathbb{N}_f \times \mathbb{N}_f$, $(a, b)\ \mathcal{R}\ (a', b')$ ssi $a \oplus b' = a' \oplus b$, où l'addition $\oplus$ est définie sur $\mathbb{N}_f$. L'ensemble des entiers relatifs graduels $\mathbb{Z}_f$ est défini comme l'ensemble quotient $(\mathbb{N}_f \times \mathbb{N}_f) / \mathcal{R}$ des classes d'équivalence sur $(\mathbb{N}_f \times \mathbb{N}_f)$ par $\mathcal{R}$.

Un entier relatif graduel x s'identifie donc à une classe d'équivalence, notée (x^+, x^-), où x^+ et x^- sont des entiers positifs graduels. Par abus de langage, nous écrirons parfois « l'entier relatif graduel (x^+, x^-) ».

Exemple 2.3.

Soit deux entiers graduels positifs :

$a = \{1/0, 1/1, 0.8/2, 0.5/3, 0.2/4\}$;

$b = \{1/0, 1/1, 0.3/2\}$.

Le couple $(a, b) = (\{1/0, 1/1, 0.8/2, 0.5/3, 0.2/4\}, \{1/0, 1/1, 0.3/2\})$ est une instance de la classe d'équivalence définissant un entier relatif graduel ($\mathbb{Z}_f$). D'autres instances de cette classe peuvent être définies par les couples suivants :

$(a', b') = (\{1/0, 0.8/1, 0.5/2, 0.2/3\}, \{1/0, 0.3/1\})$;

$(a'', b'') = (\{1/0, 1/1, 0.9/2, 0.8/3, 0.5/4, 0.2/5\}, \{1/0, 1/1, 0.9/2, 0.3/3\})$.

♦

Une entier relatif graduel $x = (x^+, x^-)$ peut donc prendre plusieurs formes équivalentes. Cependant, x peut également se décrire de façon compacte en énumérant les valeurs $(x^+_\alpha - x^-_\alpha)$ pour toutes ses α-coupes différentes. Ces valeurs sont des entiers relatifs ($\mathbb{Z}$). On obtient ainsi une représentation canonique, notée x^c, définie par :

$$x^c = \sum \alpha_i / (x^+_{\alpha_i} - x^-_{\alpha_i})$$

où les α_i correspondent aux différentes α-coupes de x et où les $(x^+_{\alpha_i} - x^-_{\alpha_i})$ appartiennent à $\mathbb{Z}$.

Exemple 2.4.

La représentation canonique de (a, b) (cf. exemple 2.3) est obtenue en énumérant les valeurs de ses différentes α-coupes :

$(a, b)^c = \{1/0, 0.8/1, 0.5/2, 0.3/1, 0.2/2\}^c$.

♦

Dans un but de simplification des notations, un nombre graduel sera décrit par la suite de manière compacte en énumérant les valeurs de ses différentes α-coupes.

Si x et y sont deux entiers relatifs graduels, l'addition et la multiplication sont respectivement définies[1] par :

$x \oplus y = (x^+ \oplus y^+, x^- \oplus y^-)$;

$x \otimes y = ((x^+ \otimes y^+) \oplus (x^- \otimes y^-), (x^+ \otimes y^-) \oplus (x^- \otimes y^+))$.

Comme dans $\mathbb{N}_f$ les opérations arithmétiques respectent la propriété $(x * y)_\alpha = x_\alpha * y_\alpha$, celle-ci est également vérifiée dans $\mathbb{Z}_f$. D'un point de vue pratique cela permet d'évaluer rapidement une opération sur $\mathbb{Z}_f$, en travaillant sur les différentes α-coupes de ses opérandes.

On note que tout entier relatif graduel (x^+, x^-) a un opposé (x^-, x^+) ce qui n'est pas vrai pour des nombres flous [10]. La différence de deux nombres relatifs graduel x et y se définit donc comme la somme de x et de l'opposé de y.

2.3 $\mathbb{Q}_f$, l'ensemble des nombres rationnels graduels

Dans cette section, nous présentons sommairement comment a été construit $\mathbb{Q}_f$, l'ensemble des nombres rationnels graduels. La démarche suivie est similaire à celle utilisée pour construire $\mathbb{Z}_f$.

Soit $\mathbb{Z}_f^*$ comme l'ensemble des éléments x de $\mathbb{Z}_f$ tel que : $\forall \alpha \in]0, 1]$, $x_\alpha \neq 0$, et $\mathcal{R}'$ la relation d'équivalence telle que :

$\forall (a, b)$ et $(a', b') \in \mathbb{Z}_f \times \mathbb{Z}_f^*$, $[a, b] \mathcal{R}' [a', b']$ ssi $a \otimes b' = a' \otimes b$.

L'ensemble des nombres rationnels graduels, noté $\mathbb{Q}_f$, est l'ensemble quotient $(\mathbb{Z}_f \times \mathbb{Z}_f^*) / \mathcal{R}'$ des classes d'équivalence sur $(\mathbb{Z}_f \times \mathbb{Z}_f^*)$ par la relation $\mathcal{R}'$.

---

[1] Dans cet article nous adoptons un principe de surcharge des opérateurs, ils se notent de la même manière dans $\mathbb{N}_f$, $\mathbb{Z}_f$ ou $\mathbb{Q}_f$.

Un nombre rationnel graduel x est donc une classe d'équivalence dont une instance peut se noter $[x^n, x^d]$ où x^n et x^d appartiennent respectivement à $\mathbb{Z}_f$ et $\mathbb{Z}_f^*$. Une telle représentation peut également s'écrire à l'aide d'entiers positifs graduels : $[(x^{n+}, x^{n-}), (x^{d+}, x^{d-})]$. Comme pour les entiers relatifs graduels, il est possible de représenter un élément de $\mathbb{Q}_f$ sous une forme compacte obtenue en énumérant les valeurs associées à ses différentes α-coupes. Ces valeurs correspondent à des nombres rationnels. La forme compacte est canonique si les rationnels de ses différentes α-coupes sont réduits.

Si x et y sont deux nombres rationnels graduels, l'addition et la multiplication sont définies par :

$$x \oplus y = [(x^n \otimes y^d) \oplus (y^n \otimes x^d), x^d \otimes y^d] ;$$

$$x \otimes y = [x^n \otimes y^n, x^d \otimes y^d].$$

D'un point de vue calculatoire, on note que l'addition et la multiplication peuvent être évaluées en les appliquant sur les différentes α-coupes de leurs opérandes.

Si dans $\mathbb{Z}_f$ tout entier a un opposé, de même, dans $\mathbb{Q}_f$, tout rationnel $[x^n, x^d]$ a un inverse $[x^d, x^n]$ (si $x^n \in \mathbb{Z}_f^*$,). L'opération de division entre deux rationnels graduels x et y, est alors définie comme la multiplication de x par l'inverse de y.

3 Relation d'ordre sur des quantités graduelles

On s'intéresse maintenant à la comparaison de quantités graduelles. Comme une quantité graduelle peut prendre plusieurs formes équivalentes, la comparaison de deux quantités graduelles peut poser quelques difficultés. Ce problème est levé en opérant sur des formes canoniques. Par ailleurs, comme $\mathbb{N}_f$ est inclus dans $\mathbb{Z}_f$ qui lui-même est un sous-ensemble de $\mathbb{Q}_f$, les relations d'ordre sur les rationnels graduelles sont également des relations d'ordre sur les entiers graduels.

Une relation d'ordre partielle booléenne entre deux quantités graduelles peut être définie par :

$$\forall (a, b) \in \mathbb{Q}_f,\ a \geq b \text{ ssi } a \ominus b \in \mathbb{Q}_f^{\ +},$$

où $\ominus$ est la différence sur $\mathbb{Q}_f$ et $\mathbb{Q}_f^+$ est l'ensemble des entiers relatifs graduels dont toutes les α-coupes sont positives ou nulles. Autrement dit :

$$\forall\ \alpha \in\ ]0, 1],\ a_\alpha - b_\alpha \geq 0,$$

où - est la différence sur $\mathbb{Q}$; les valeurs associées à la forme canonique $(a \ominus b)^c$ sont donc positives ou nulles. Cette relation d'ordre est partielle car il suffit que pour une coupe α la propriété $a_\alpha - b_\alpha \geq 0$ ne soit pas satisfaite, pour que a et b ne soient pas comparables au sens de la relation ≥ (même si α est faible).

Nous proposons, dans un premier temps, une extension de la relation d'ordre à valeur booléenne définie en définissant une relation d'ordre graduelle (*i.e.* à valeur dans [0, 1]), notée $\geq_g$, exprimant que plus un grand nombre d'α-coupes de niveau élevé satisfont la propriété $a_\alpha - b_\alpha \geq 0$, plus, globalement, $a \geq_g b$ est satisfait.

Puis, dans un second temps, nous examinons comment introduire une notion de tolérance floue au sein même de la relation d'ordre afin de permettre à certaines α-coupes d'enfreindre, quelque peu, la relation d'ordre. Cette relation d'ordre graduelle et tolérante est notée $\geq_{gT.}$

3.1 *Relation d'ordre graduelle*

Soit les deux quantités graduelles a et b suivantes :

$$a = \{1/0, 0.8/1, 0.5/2\} ;$$

$$b = \{1/0, 0.9/1, 0.4/2\}.$$

Leur différence $delta = a \ominus b$ s'exprime sous une forme canonique par :

$$delta = \{1/0, 0.9/\text{-}1, 0.8/0, 0.5/1, 0.4/0\}^c$$

et peut se représenter graphiquement par :

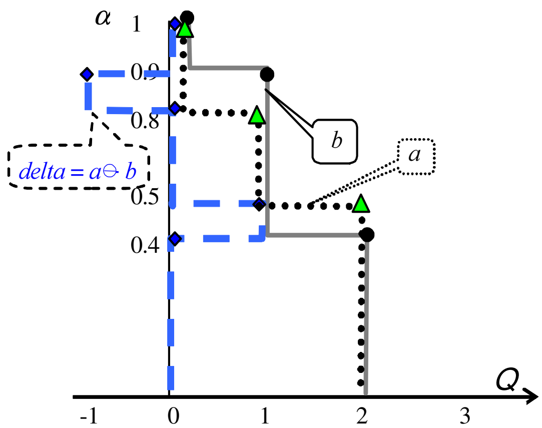


Figure 1. *Représentation graphique de deux entiers graduels a et b et de leur différence*

La satisfaction globale de la propriété $a \geq b$, notée S, établissant un bilan α-coupe par α-coupe de la satisfaction de a_α - $b_\alpha \geq 0$, peut se schématiser par la figure 2.

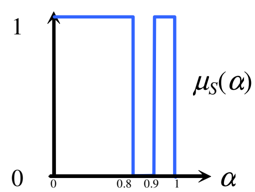


Figure 2. *Satisfaction globale S de $a \geq b$*

Un tel résultat est un ensemble S qui se définit par :

$\mu_S(\alpha) = 0 \Leftrightarrow (delta_\alpha < 0)$;

$\mu_S(\alpha) = 1 \Leftrightarrow (delta_\alpha \geq 0)$.

La question qui se pose est alors de « résumer » une satisfaction globale par un degré entre 0 et 1, c'est-à-dire d'en trouver une mesure M traduisant la sémantique « plus il y a un grand nombre d'α-coupes de niveau élevé satisfaisant la propriété a_α - $b_\alpha \geq 0$, plus, globalement, $a \geq b$ est satisfait ». Une telle mesure définit une relation d'ordre graduelle notée $\geq_g$: $(a \geq_g b) = M(S)$ qui doit respecter un certain nombre de propriétés que nous analysons ci-dessous. Nous montrons ensuite qu'une telle relation d'ordre graduelle généralise une implication floue.

3.1.1 *Définition d'une mesure de satisfaction globale relative à la relation $\geq_g$*

Soit deux nombres rationnels graduels a et b. L'évaluation de la relation d'ordre $a \geq_g b$ conduit à établir une mesure M de la satisfaction globale établissant un bilan des satisfactions de a_α - $b_\alpha \geq 0$ pour α compris entre 0 et 1. Avant de proposer une définition pour M, examinons les propriétés attendues pour une telle mesure M.

Propriété 1. Lorsque que a et b sont tels que la propriété a_α - $b_\alpha \geq 0$ est satisfaite pour tout α, la relation d'ordre est entièrement satisfaite. La satisfaction globale de la propriété $a \geq_g b$ est l'ensemble flou noté S_1 tel que : $\forall\ \alpha \in\ ]0, 1]$, $\mu_{S_1}(\alpha) = 1$. La mesure M de cet ensemble, qui évalue la satisfaction du prédicat graduel $(a \geq_g b)$, vaut 1.

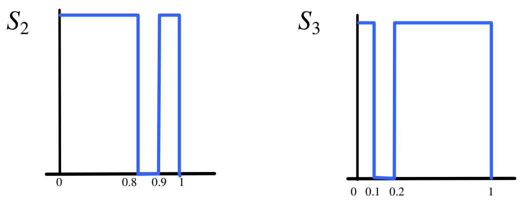
Propriété 2. Lorsqu'aucune α-coupe ne satisfait la relation d'ordre, l'ensemble flou représentant la satisfaction globale est $S_0 = \varnothing$ ($\forall\ \alpha \in\ ]0, 1]$, $\mu_{S_0}(\alpha) = 0$) et sa mesure M vaut 0.

Propriété 3. Lorsque a et b sont tels que seules les α-coupes entre 0 et un seuil t satisfont la propriété a_α - $b_\alpha \geq 0$, la satisfaction globale est l'ensemble flou S_t tel que : $\mu_{S_t}(\alpha) = 1$ si $\alpha \in\ ]0, t]$, $\mu_{S_t}(\alpha) = 0$ sinon. La mesure M de S_t vaut t, elle exprime que le prédicat est complètement satisfait jusqu'au degré t.

Cette propriété 3 est cohérente avec les deux premières.

Propriété 4. Lorsque deux satisfactions globales S_2 et S_3 ont le même nombre d'α-coupes satisfaites, la mesure de S_3 est supérieure à celle de S_2, si S_3 contient un plus grand nombre d'α-coupes de haut niveau satisfaites que S_2.

Considérons les deux ensembles flous correspondant aux évaluations globales S_2 et S_3 de $a_2 \geq b_2$ et $a_3 \geq b_3$ suivants :



$M(S_2) \quad \leq \quad M(S_3)$

Figure 3. Comparaison de deux satisfactions globales

On constate qu'il y a le même nombre d'α-coupes satisfaisantes dans S_2 et dans S_3. Cependant, dans S_2 les α-coupes ayant une satisfaction nulle sont de degré plus élevé que celles de S_3. Ainsi, globalement, on peut dire que la relation d'ordre est mieux satisfaite par a_3 et b_3 que par a_2 et b_2 d'où $M(S_2) \leq M(S_3)$. Autrement dit, plus une α-coupe est de niveau élevé, plus son degré de satisfaction contribue à l'évaluation de M.

Propriété 5. Si S_4 (respectivement S_5) est l'ensemble flou représentant la satisfaction globale de la propriété $a_4 \geq b_4$ (resp. $a_5 \geq b_5$) et que l'ensemble flou S_4 est inclus dans S_5, nous déduisons qu'un plus grand nombre d'α-coupes de S_5 satisfont la relation d'ordre. Dans ce cas $M(S_4) \leq M(S_5)$.

Différentes mesures M prenant en compte tout ou partie des propriétés précédemment énoncées peuvent être envisagées. Nous construisons maintenant progressivement une mesure satisfaisant nos objectifs.

La mesure M_1 définie comme l'intégrale de la fonction caractéristique $\mu_S(\alpha)$ de l'ensemble flou des satisfactions par α-coupes :

$$M_1(S) = \int_0^1 \mu_S(\alpha)\, d\alpha$$

satisfait les propriétés 1, 2, 3 et 5 mais ne tient pas compte du niveau des α-coupes satisfaisantes (*i.e.* la propriété 4 n'est pas satisfaite). Pour ce faire, il faut pondérer $\mu_S(\alpha)$ par une fonction de pondération croissante, par exemple par 2α, soit :

$$M'_2(S) = \int_0^1 \mu_S(\alpha)\; 2\alpha\; d\alpha .$$

Dans ce cas, les propriétés 1, 2, 4, 5 sont satisfaites mais la propriété 3 ne l'est plus. En effet, si la satisfaction globale est l'ensemble flou S tel que $\forall \alpha \in \,]0, t]$, $\mu_S(\alpha) = 1$ et $\forall \alpha \in \,]t, 1]$, $\mu_S(\alpha) = 0$, alors :

$$M'_2(S) = \int_0^t 2\alpha\; d\alpha = \left[\alpha^2\right]_0^t = t^2 .$$

Il en découle la mesure M_2 normalisée suivante, satisfaisant les propriétés 1 à 5 :

$$M_2(S) = \left(\int_0^t 2\alpha\; d\alpha\right)^{1/2}$$

Ainsi, pour les différents cas présentés précédemment (propriétés 1 à 5), nous obtenons les évaluations suivantes :

$$M_2(S_1) = \left(\int_0^1 2\alpha\; d\alpha\right)^{1/2} = 1\,;$$

$$M_2(S_0) = \left(\int_0^1 0\; d\alpha\right)^{1/2} = 0\,;$$

$$M_2(S_2) = \left(\int_0^{0.8} \mu_{S_2}(\alpha)\; 2\alpha\; d\alpha + \int_{0.9}^1 \mu_{S2}(\alpha)\; 2\alpha\; d\alpha\right)^{1/2}$$

$$= \left(\left[\alpha^2\right]_0^{0.8} + \left[\alpha^2\right]_{0.9}^{1}\right)^{1/2} = 0.91\;;$$

$$M_2(S_3) = \left(\int_0^{0.1} \mu_{S_3}(\alpha)\; 2\alpha\; d\alpha + \int_{0.2}^1 \mu_{S_3}(\alpha)\, 2\alpha\; d\alpha\right)^{1/2}$$

$$= \left(\left[\alpha^2\right]_0^{0.1} + \left[\alpha^2\right]_{0.2}^{1}\right)^{1/2} = 0.98 .$$

Ces différentes mesures sont bien ordonnées conformément à l'ordre induit par la sémantique « plus il y a un grand nombre d'α-coupes de niveau élevé satisfaites, plus la relation d'ordre est satisfaite ».

3.1.2 *Relation d'ordre graduelle et implication floue*

La relation d'ordre graduelle $\geq_g$ possède la particularité de généraliser les implications floues qui, elles-mêmes, généralisent la notion d'implication usuelle en satisfaisant un certain nombre d'axiomes comme la monotonie décroissante (resp. croissante) par rapport au premier (resp. second) argument, $(0 \Rightarrow_f a) = 1$, $(1 \Rightarrow_f 1) = 1$ ou $(1 \Rightarrow_f a) = a$. Par exemple, les trois implications floues suivantes sont couramment utilisées :

- l'implication de Lukasiewicz (1 si $a \leq b$, 1 – $a + b$ sinon) ;
- l'implication de Godël (1 si $a \leq b$, b sinon) ;
- l'implication de Goguen (1 si $a \leq b$, b/a sinon).

L'idée développée ici est que, de même que la comparaison de deux degrés peut être interprétée comme une implication floue, la comparaison de deux entiers graduels peut être interprétée comme une implication floue généralisée. Une telle implication floue généralisée se réduit donc à une implication floue dans le cas particulier où les entiers graduels considérés représentent des degrés (un entier graduel peut être vu comme un degré α lorsqu'il prend la forme particulière $\{1/0, \alpha/1\}$).

Dans ce cas, évaluons la condition $\{1/0, a/1\} \leq \{1/0, b/1\}$ et vérifions la compatibilité du résultat avec une implication floue $a \Rightarrow_f b$.

Pour cela, on calcule la différence *delta* = $\{1/0, b/1\} \ominus \{1/0, a/1\}$.

Si $a \leq$ b, delta est l'entier relatif graduel noté en représentation canonique : $\{1/0, b/1, a/0\}^c$. Toutes ses coupes appartiennent à $\mathbb{Z}^+$, on en déduit que la mesure de la satisfaction globale du prédicat (*delta* ≥ 0) est 1, ce qui est cohérent avec l'évaluation de $a \Rightarrow_f b$.

Si $a > b$, *delta* est l'entier relatif graduel dont la représentation canonique est $\{1/0, a/\text{-}1, b/0\}^c$. La satisfaction globale de $\{1/0, a/1\} \leq \{1/0, b/1\}$ (ou *delta* ≥ 0) peut se représenter graphiquement par l'ensemble flou S de la figure 4.

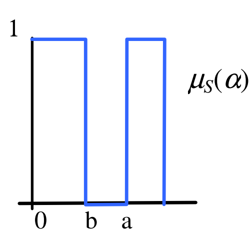


Figure 4. *Satisfaction globale de* $a \Rightarrow_f b$

La mesure $M_1(S) = 1 - a + b$ est homogène avec le résultat obtenu par l'évaluation de $a \Rightarrow_f$ b, avec $\Rightarrow_f$ l'implication floue de Lukasiewicz. La mesure $M_2(S) = (1 - a^2 + b^2)^{1/2}$ est, quant à elle, homogène avec l'implication dite quadratique, variante de l'implication de Lukasiewicz.

De façon plus générale, une classe importante d'implications floues, dites de Schweizer-Sklar, se définit par $(1\text{-}\ a^p + b^p)^{1/p}$ quand $a > b$, 1 sinon [31]. Dans le cas limite de p tendant vers $+\infty$, on retrouve l'implication de Godël. Quand p tend vers 0, on retrouve l'implication dite de Goguen. La méthode que nous avons présentée nous permet de retrouver ces résultats et en propose une généralisation. En effet, il suffit de choisir une intégrale pondérée par la fonction croissante $p.\alpha^{p-1}$, ce qui établit la mesure M_p' :

$$M_p'(S) = \int_0^1 \mu_S(\alpha)\ p\alpha^{p-1}\ d\alpha$$

puis de définir $M_p = (M_p')^{1/p}$ afin de normaliser le résultat et satisfaire la propriété 3. Cette dernière propriété correspond, en fait, à l'axiome des implications : $(1 \Rightarrow_f b) = b$.

On notera, sans le développer ici, qu'une expression des implications floues encore plus générale [32] aurait pu être retrouvée en utilisant une fonction de pondération croissante $\varphi(\alpha)$ quelconque et son inverse φ^{-1} à des fins de normalisation.

Cette présentation d'une relation d'ordre sur des quantités graduelles vue comme une implication floue généralisée est importante car elle s'inscrit dans notre démarche de construction du concept de multi-ensemble flou comme généralisation à la fois du concept d'ensemble flou et du concept de multi-ensemble. En effet, les opérateurs sur les ensembles flous, basés sur des opérations sur des degrés, et les opérateurs sur les multi-ensembles, basés sur des opérations sur des entiers, peuvent être vus comme des cas particuliers d'opérateurs sur les multi-ensembles flous, basés sur des entiers graduels. De manière analogue, la relation d'ordre sur des entiers graduels, sous-jacente à une inclusion dans le cadre des multi-ensembles flous, généralise, d'une part, une implication entre degrés et, d'autre part, une relation d'ordre sur les entiers. Le lien entre la relation d'ordre sur les entiers et celle sur les entiers graduels est assez facile à établir et nous ne le développerons pas dans ce papier.

3.2 Relation d'ordre graduelle tolérante

La relation d'ordre $\geq_g$, définie en section 3.1, rend graduelle une relation d'ordre à valeur booléenne entre quantités graduelles et, de ce fait, définit une relation d'ordre total. Nous étudions maintenant une seconde extension, notée $\geq_{gT}$, consistant en une relaxation floue de la relation d'ordre graduelle $\geq_g$.

Nous souhaitons comparer deux quantités graduelles a et b ($a \geq_{gT} b$) en introduisant une idée de tolérance graduelle dans la mesure des écarts $delta_\alpha = a_\alpha - b_\alpha$. L'objectif est d'autoriser, dans une certaine mesure, des écarts légèrement négatifs pour certaines α-coupes. La relation d'ordre tolérante ainsi construite est alors mieux satisfaite que la relation d'ordre graduelle qu'elle relaxe. Cette tolérance peut s'exprimer au moyen d'un ensemble flou T, par exemple l'ensemble flou "au moins 0" représenté par la figure 5.

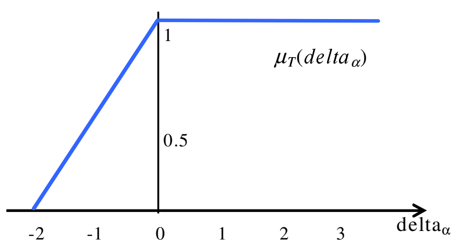


Figure 5. Prédicat au moins 0 avec tolérance

Alors que dans le cas de la relation graduelle $\geq_g$ la satisfaction de chaque $delta_\alpha$ était soit 0, soit 1, celle-ci est maintenant comprise entre 0 et 1, compte tenu de la tolérance T. Pour les quantités graduelles décrites par la figure 1 et de la tolérance T décrite par la figure 5, la satisfaction globale S_T de $a \geq_{gT} b$, est définie par $\mu_{S_T}(\alpha) = \mu_T(delta_\alpha)$, elle se représente graphiquement par :

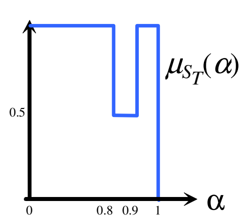


Figure 6. *S_T la satisfaction globale de $a \geq_{gT} b$*

De la même manière que précédemment, la mesure de cet ensemble peut s'exprimer en évaluant une intégrale pondérée, ce qui conduit à évaluer le degré satisfaction de $a \geq_{gT} b$ par :

$$M_p(S_T) = (\int_0^1 \mu_{S_T}(\alpha)\ p\alpha^{p-1}\ d\alpha)^{1/p}.$$

En choisissant $p = 2$, correspondant à une implication quadratique, on obtient : $M_2(S_T) = 0.95$.

3.3 Requêtes flexibles comportant une comparaison entre quantités graduelles

L'utilisation de quantités graduelles et de relations d'ordre sur ces quantités, rend possible le traitement de requêtes flexibles sur une base de données usuelle comportant des prédicats sur des cardinalités d'ensembles flous. À titre d'exemple nous traitons les trois requêtes simples suivantes :

1) *trouver les entreprises dont le nombre des employés jeunes est supérieur à celui des employés proches de la retraite* ;
2) *trouver les entreprises ayant plus de deux employés jeunes et bien payés* ;
3) *trouver les entreprises dont la plupart des employés jeunes sont bien payés.*

Exemple 3.1.

Pour traiter la première requête, supposons qu'une entreprise E de la base de données ait dix employés e_1, …, e_{10} dont les âges respectifs sont : 25, 30, 35, 40, 42, 43, 48, 48, 55, 60 et que les concepts *employé jeune* et *proche de la retraite* soient définis par les ensembles flous sur les âges décrits par la figure 7.

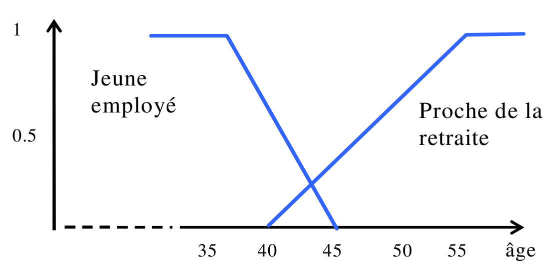


Figure 7. *Les prédicats flous jeune et proche de la retraite*

Compte tenu de ces prédicats, l'ensemble flou des employés jeunes correspond à $\{1/e_1,\ 1/e_2,\ 1/e_3,\ 0.5/e_4,\ 0.3/e_5,\ 0.2/e_6\}$ et l'ensemble flou des employés proches de la retraite est $\{0.15/e_5,\ 0.2/e_6,\ 0.55/e_7,\ 0.55/e_8,\ 1/e_9,\ 1/e_{10}\}$.

Les cardinalités respectives de ces ensembles flous sont : {1/0, 1/2, 1/3, 0.5/4, 0.3/5, 0.2/6} et {1/0, 1/2, 0.55/3, 0.55/4, 0.2/5, 0.15/6}.

Pour comparer ces deux nombres on évalue leur différence qui est : $\{1/1,\ 0.55/\text{-}1,\ 0{,}5/0,\ 0.3/1,\ 0.2/0\}^c$. Ce résultat signifie que le prédicat c = *nombre d'employés jeunes* $\geq_g$ *nombre d'employés proches de la retraite* est complètement satisfait si l'on considère des α-coupes appartenant à]0, 0.5] ∪]0.55, 1], ce qui définit la satisfaction globale S_c du prédicat p.

La mesure M_p de cette satisfaction globale, en considérant que plus les employés sont jeunes ou proches de la retraite plus il est important que le nombre des premiers soit supérieur au nombre des seconds est (en utilisant $p = 2$ pour exprimé la pondération):

$$M_2(S_c) = (\int_0^1 \mu_{S_c}(\alpha)\ 2\alpha\ d\alpha)^{1/2}$$
$$= \sqrt{0.25 + 0.69} = 0.97$$

On détermine ainsi dans quelle mesure l'entreprise E considérée a plus d'employés jeunes que d'employés proches de la retraite. De cette manière, à partir d'une base d'entreprises, on peut sélectionner celles satisfaisant plus ou moins le critère de sélection. Les entreprises retournées peuvent alors être ordonnées par rapport à leur niveau d'adéquation au critère.

Exemple 3.2.

La seconde requête *trouver les entreprises ayant plus de deux employés jeunes et bien payés* comporte un quantificateur absolu *plus de 2*. Elle se traite simplement en comparant la cardinalité floue n de l'ensemble flou des employées *jeunes et bien payés* (qui sous forme de nombre graduel s'écrit {1/0, 1/1, 1/2}) avec la quantité 2. Cette comparaison conduit à réaliser la différence entre n et la valeur 2, dans $\mathbb{Z}_f$, puis à appliquer M_p évaluant dans quelle mesure le nombre obtenu est positif.

Ainsi, soit l'ensemble flou des personnes *jeunes* $\{1/p_1,\ 1/p_2,\ 0.4/p_3,\ 0.1/p_4\}$ d'une entreprise donnée E. Supposons que ces personnes aient les salaires respectifs 2500, 1000, 2000 et 1000 €, et que $\mu_{bienPayé}(2500) = 1$, $\mu_{bienPayé}(2000) = 0.7$ et $\mu_{bienPayé}(1000) = 0.2$. Nous en déduisons l'ensemble des personnes *jeunes et bien-payées* = $\{1/p_1,\ 0.2/p_2,\ 0.4/p_3,\ 0.1/p_4\}$ dont la cardinalité n est notée : {1/1, 0.4/2, 0.2/3, 0.1/4}.

La différence entre cette cardinalité et 2 est l'entier relatif graduel : $\{1/\text{-}1,\ 0.4/0,\ 0.2/1,\ 0.1/2\}^c$. La satisfaction globale S_c du prédicat c = "$n \geq_g 2$" est donc telle que : $\mu_{Sc}(\alpha) = 1$ si $\alpha \in\ ]0, 0.4]$, $\mu_{Sc}(\alpha) = 0$ sinon.

La mesure $M_p(S_c)$ est égale à 0.4 (du fait de la propriété 3), ce qui est bien conforme à la sémantique de la cardinalité n où le degré 0.4 exprime dans quelle mesure l'ensemble flou dont la cardinalité est n contient au moins 2 éléments.

Relaxons maintenant quelque peu la relation d'ordre $x \geq_g 2$, en définissant une relation d'ordre tolérante $x \geq_{gT} 2$ traduisant l'expression « x est à peu près plus grand que 2 ». Pour ce faire, on utilise, par exemple, la tolérance définie par la figure 5. Dans ce cas, la satisfaction du prédicat c' = "$n \geq_{gT} 2$" est l'ensemble flou $S_{c'}$ défini par : $\mu_{Sc'}(\alpha) = 1$ si $\alpha \in\ ]0, 0.4]$, $\mu_{Sc'}(\alpha) = 0.5$ sinon. La mesure M_p de $S_{c'}$ est (en choisissant, par exemple, $p = 1$) :

$$M_1(S_{c'}) = \int_0^1 \mu_{Sc'}(\alpha)\ d\alpha = 0.4 + 0.3 = 0.7.$$

Cette mesure établit un bilan des satisfactions par α-coupes du prédicat c', en considérant que chaque α-coupe a la même importance. On détermine ainsi dans quelle mesure une entreprise E considérée a à peu près plus de 2 employés jeunes et bien payés.

Exemple 3.3.

On s'intéresse maintenant à la *requête trouver les entreprises dont la plupart des employés jeunes sont bien payés.* Cette requête se traite en évaluant la proportion du nombre d'employés *jeunes et bien payés* par rapport au nombre d'employés *jeunes*. Plus ce rapport est proche de 1, plus la plupart des employés jeunes sont bien payés. Quand le rapport vaut 1, tous les employés jeunes sont bien payés.

On réalise donc, dans $\mathbb{Q}_f$, la division entre deux cardinalités floues. Le quotient obtenu est ensuite comparé à 1, c'est-à-dire que la quantité graduel quotient est filtrée par un prédicat flou "presque 1" préalablement spécifié, comme :

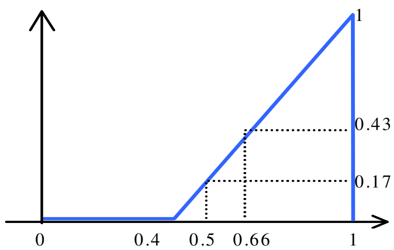


Figure 8. Prédicat flou presque 1

Ainsi, si la cardinalité de l'ensemble flou des employés *jeunes* est $\{1/2, 0.4/3, 0.1/4\}^c$, le rapport du nombre d'employés *jeunes et bien payés* sur le nombre d'employés *jeunes* est le nombre rationnel graduel représenté sous forme compacte par : $\{1/1{:}2, 0.4/2{:}3, 0.2/1\}^c$.

En évaluant ce rapport en fonction du prédicat presque 1, nous en déduisons une satisfaction globale S qui est représentée graphiquement par :

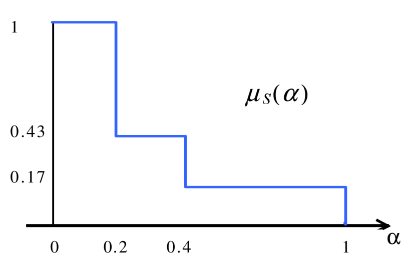


Figure 9. *Presque 1 sur $\{1/1{:}2, 0.4/2{:}3, 0.2/1\}^c$*

Cette satisfaction peut se résumer par $M_2(S)$:

$$M_2(S) = (\int_0^1 \mu_S(\alpha) * 2\alpha * d\alpha)^{1/2} = \sqrt{0.23} = 0.48.$$

4 Expressions quantifiées

Les nombres graduels ($\mathbb{N}_f$, $\mathbb{Z}_f$ et $\mathbb{Q}_f$) que nous avons présentés en sections 2 et 3 offrent un cadre général pour évaluer des expressions quantifiées. Les quantificateurs flous ont été introduits par Zadeh afin de généraliser les quantificateurs universel et existentiel usuels. On distingue les quantificateurs absolus exprimant un nombre (*environ 3*, *au moins 2*) ou relatifs référençant une proportion (*la plupart*, *environ la moitié*). Les expressions quantifiées sont de type "*Q X* sont *A*" lorsqu'il s'agit d'évaluer dans quelle mesure *Q éléments de X sont A*, comme dans "environ 3 employés sont bien payés". Elles se généralisent en "*Q B X* sont *A*" où l'ensemble de référence *B X* est flou comme dans "la plupart des employés jeunes sont bien payés" (où *B* est le prédicat flou jeune).

L'évaluation d'expressions quantifiées a donné lieu à de nombreuses propositions. On en trouve des études détaillées dans [11][20]. Les plus importantes sont celles basées sur l'opérateur OWA[2] [35] et l'intégrale de Sugeno [5]. Ces deux approches délivrent un degré de satisfaction mais sont limitées à des expressions utilisant des quantificateurs croissants. Une approche basée sur la notion de nombres graduels et de relations d'ordre graduelles, n'impose pas d'hypothèse particulière sur la monotonie du quantificateur linguistique et généralise l'approche avec des OWA. Qui plus est, cette méthode d'évaluation permet également d'englober d'autres approches.

Dans la suite nous rappelons d'abord la notion de quantificateur et l'interprétation d'une expression quantifiée au moyen d'un opérateur OWA. Puis nous montrons comment les outils relatifs aux nombres graduels permettent de fournir une interprétation assez simple et naturelle des expressions comportant un quantificateur absolu ou relatif.

4.1 Quantificateurs linguistiques et approche OWA

Deux sortes de quantificateurs flous peuvent être distinguées : les quantificateurs absolus, comme *environ 2*, *au moins 3*…, et les quantificateurs relatifs se référant à des proportions, comme *environ la moitié*, *au moins un quart*… Ces quantificateurs peuvent être croissants (resp. décroissants) ce qui signifie qu'un accroissement de satisfaction de la condition *A* ne peut faire décroître (croitre) la satisfaction de la proposition "*Q X* sont *A*". *Au moins 3* et *presque tous* sont des

[2] Ordered Weighted Average

exemples de quantificateurs croissants. Un quantificateur est monotone s'il est soit croissant, soit décroissant. On trouve également des quantificateurs unimodaux qui font référence à des prédicats graduels comme *environ 4* ou *environ la moitié*. Les quantificateurs absolus sont modélisés par des ensembles flous sur les réels (voir figure 10) alors que les quantificateurs relatifs correspondent à des ensembles flous sur [0, 1].

L'interprétation de "Q X sont A", avec Q croissant, au moyen d'un opérateur OWA [35] est donnée par :

$$\text{OWA} = \sum_{i=1}^{n} w_i \times \mu_A(x_i),$$

où les degrés d'appartenance à A sont ordonnés de manière décroissante ($\mu_A(x_1) \geq \mu_A(x_2) \geq \ldots \geq \mu_A(x_n)$). Pour un quantificateur absolu les poids $w_i = \mu_Q(i) - \mu_Q(i\text{-}1)$, expriment l'accroissement de satisfaction lorsqu'on compare une situation comportant i-1 éléments entièrement A avec une situation comportant i éléments entièrement A. Pour un quantificateur relatif, les poids $w_i = \mu_Q(i/n) - \mu_Q((i\text{-}1)/n)$, avec n la cardinalité de l'ensemble X, expriment l'accroissement de satisfaction de la quantification lorsque la proportion des éléments de X satisfaisant complètement A passe de i-1/n à i/n.

Exemple 4.1.

Q est le quantificateur absolu *au moins 3*, défini par la figure 5, et $X = \{x_1, x_2, x_3, x_4\}$ avec $A = \{1/x_1, 1/x_2, 0.8/x_3, 0.6/x_4\}$.

Il en découle que les poids w_i sont :

$w_1 = 0$, $w_2 = 0.5$, $w_3 = 0.5$, $w_4 = 0$,

et que la satisfaction de la proposition "au moins 3 X sont A" est :

$$\text{OWA} = 0 \times 1 + 0.5 \times 1 + 0.5 \times 0.8 + 0 \times 0.6 = 0.9.$$

Ce degré proche de 1 est conforme à l'intuition puisque au moins 3 éléments de A ont un degré élevé.

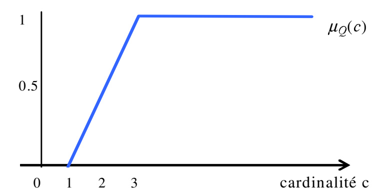


Figure 10. *Le quantificateur au moins 3.*

♦

4.2 Interprétation d'une expression quantifiée basée sur des quantités graduelles

L'interprétation d'une expression quantifiée, comme "au moins 3 employés sont bien payés", peut s'appuyer sur une approche basée sur les nombres graduels. Il suffit de considérer la cardinalité floue de l'ensemble des employés bien payés (notée $|E_{bp}|$) et de la comparer à l'entier 3 au moyen d'une relation d'ordre graduelle associée à une fonction de tolérance T (comme celle décrite par la figure 3). On évalue donc : $|E_{bp}| \geq_{gT} 3$.

Selon la méthode définie en section 3, cela consiste à filtrer les différentes α-coupes de $|E_{bp}| - 3$ au moyen de la fonction T. Le résultat est alors une satisfaction globale par α-coupes S dont il est ensuite possible d'évaluer différents résumés en appliquant une mesure M_p. Si on choisit $p = 1$, $M_1(S)$ correspond au résultat délivré par l'opérateur OWA.

Exemple 4.2.

Reprenons les données de l'exemple 4.1. L'évaluation de l'expression « au moins 3 X sont A », en considérant la fonction de tolérance de la figure 5, conduit à la satisfaction globale décrite par la figure 11.

En appliquant une mesure M_1 sur la satisfaction globale S, on obtient un degré qui peut être vu comme un résumé de cette expression, complexe mais exacte, de la satisfaction de la quantification considérée. L'utilisation d'une mesure M_1 conduit à interpréter la *relation d'ordre graduelle tolérante* sous-jacente comme une implication floue de Lukasiewicz généralisée. Le résultat obtenu correspond à celui issu de l'application de l'opérateur OWA.

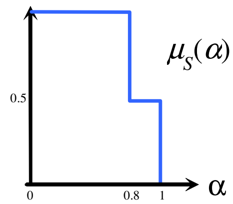


Figure 11. *Satisfaction globale S de « au moins 3 X sont A »*

♦

L'utilisation d'une mesure M_1 attribue une même importance à chaque α-coupe. Mais d'autres mesures M_p sont envisageables. Elles permettent de traduire la sémantique *plus un grand nombre d'α-coupes de haut niveau sont satisfaites plus globalement la quantification considérée est satisfaite*. Pour cela, il suffit de choisir une mesure M_p avec un paramètre p supérieur à 1, ce qui permet d'attribuer une importance plus grande aux α-coupes de degré élevé. De cette manière, on considère que les quantités issues du décompte des éléments les plus significatifs (parce que les plus satisfaisants) doivent d'autant plus respecter la relation d'ordre sous-jacente à la quantification traitée.

Cette méthode s'applique également sans difficulté pour évaluer une expression quantifiée dont le quantificateur est relatif, comme dans "plus de la moitié des employés sont bien payés". Il suffit de travailler sur $\mathbb{Q}_f$, et de calculer le nombre rationnel graduel R représentant le ratio de la cardinalité floue des employés bien payés par la cardinalité des employés, puis d'évaluer la relation $R \geq_{gT} 1$, avec une relation de tolérance T adaptée au quantificateur Q. Le bilan de cette comparaison est une satisfaction globale dont différentes approximations peuvent être envisagées par application d'une mesure M_p.

Plus généralement, l'évaluation d'une expression du type "*Q B X* sont *A*", comme "la plupart des employés jeunes sont bien payés", se fait en calculant le nombre rationnel graduel correspondant au ratio de la cardinalité floue des employés jeunes et bien payés sur la cardinalité floue des employés jeunes. Plus il y a d'employés jeunes qui sont bien payés plus ce ratio est proche de 1. La méthode employée est donc similaire à celle utilisée pour évaluer les expressions "*Q X* sont *A*".

Le point clé de la méthode proposée est d'exprimer la satisfaction d'une expression quantifiée non pas sous la forme d'un degré mais sous la forme d'une *satisfaction globale S* représentée par un ensemble flou sur [0, 1]. Cet ensemble flou décrit *précisément et complètement* la manière dont la proposition est satisfaite. Une telle satisfaction peut être vue comme une *valeur de vérité généralisée*.

Ainsi, notre méthode basée sur la comparaison de nombres graduels généralise l'approche OWA et définit une technique unifiée pour traiter des expressions quantifiées de type "*Q X* sont *A*" et "*Q B X* sont *A*".

Par ailleurs, l'approche d'évaluation d'une expression quantifiée fondée sur une intégrale de Sugeno définit le score d'une proposition "*Q X* sont *A*" par :

$$d = \max(\ \min(\mu_Q(j),\ \mu_{|A\ X|}(j)) \text{ pour } j = 0, \ldots, |X|.$$

Le degré d s'interprète comme une mesure qualitative évaluant le meilleur compromis entre la qualité du représentant choisi pour A et l'adéquation de sa cardinalité vis à vis de Q. On constate que ce degré n'est autre qu'*une approximation particulière de la satisfaction globale S* générée par l'interprétation de la proposition quantifiée au moyen d'une relation d'ordre graduelle appliquée à des nombres graduels. Cependant, avec l'approche de Sugeno, le calcul de d n'est significatif que lorsque le quantificateur est monotone croissant. Ainsi, dans [4], il a été proposé une technique d'évaluation d'une propriété graduelle sur un ensemble flou en généralisant une intégrale de Sugeno. Dans le cas d'une proposition quantifiée cette méthode permet de s'affranchir de la contrainte liée à la monotonie en tenant compte à la fois de la qualité du représentant et de sa représentativité. Cette approche traduit la sémantique que *plus il y a d'α-coupes de A satisfaisant de manière élevée le quantificateur Q, plus la proposition quantifiée est satisfaite*. On constate alors que cette démarche n'est plus complètement qualitative car elle prend en compte une notion de décompte sur les α-coupes. L'approximation calculée est donc basée sur une approche intermédiaire entre une approche quantitative et une approche qualitative (basée sur une intégrale de Sugeno).

De notre point de vue ces différentes approches ont toutes un point commun, celui d'être basées sur l'évaluation d'une comparaison entre nombres graduels. Ce calcul produit une satisfaction globale reflétant exactement la manière dont une proposition est satisfaite. Vient ensuite un processus de '*defuzzification*' où différentes techniques peuvent être utilisées, ce qui correspond à une autre problématique que nous traitons par le choix d'une mesure M_p.

En séparant les problèmes à traiter, la comparaison de quantités d'une part, la production de résumés d'autre part (différentes techniques de '*defuzzification*' pouvant être considérées), nous englobons les approches existantes dans un même processus et nous ouvrons le sujet sur de nouvelles propositions.

5 Conclusion

La démarche proposée repose sur la construction d'une structure générale, les multi-ensembles flous, intégrant à la fois les concepts d'ensemble flou et de multi-ensemble. L'originalité de cette structure provient de l'émergence du concept d'entier graduel ($\mathbb{N}_f$) qui, en étendant à la fois la notion d'entier et de valeur de vérité, se démarque de la notion de nombre flou (imprécis). Les entiers graduels offrent un unique formalisme pour caractériser les ensembles, multi-ensembles, ensembles flous, multi-ensemble flous. Dès lors, ceux-ci peuvent être traités de la même manière grâce à un faible nombre d'opérateurs algébriques génériques. Les extensions ainsi obtenues apparaissent assez "intuitives" car elles étendent "naturellement" des concepts usuels. De plus, $\mathbb{N}_f$ a pu être prolongé en $\mathbb{Z}_f$, l'ensemble des entiers relatifs graduels, puis $\mathbb{Q}_f$, l'ensemble des nombres rationnels graduels, définissant ainsi les outils de base pour traiter des notions de quantités absolues ou relatives.

Un des fils conducteurs de nos propositions a été de définir des systèmes de calculs exacts sur des quantités graduelles puis, en fonction des besoins, d'offrir à l'utilisateur différents résumés des résultats exacts (mais graduels) obtenus. Nous avons proposé d'évaluer ces résumés au moyen d'intégrales pondérées par des fonctions transformant des degrés d'appartenance en degrés d'importance, mais l'approche est ouverte et d'autres techniques peuvent être considérées.

Le cadre ainsi obtenu permet de définir une approche pour évaluer des expressions quantifiées généralisant celles basées sur l'opérateur OWA ou sur une intégrale de Sugeno.

Ce travail sur la notion de relation d'ordre peut être complété en étudiant de quelle manière des prédicats comportant des relations d'ordre sur des quantités graduelles peuvent être composés au moyen d'opérateurs conjonctifs, disjonctifs ou d'autres opérateurs d'agrégation. Autrement dit, il s'agit de définir des opérateurs logiques composant des satisfactions globales vues comme des valeurs de vérité généralisées.

Par ailleurs, dans le cadre du projet APMD (http://apmd.prism.uvsq.fr) défini au sein du programme nationale ACI Masses de Données, nous étudions actuellement les problèmes d'implémentation liés à l'évaluation de requêtes flexibles. Cet aspect important de validation de l'approche proposée repose sur la construction d'un plan d'évaluation d'une requête flexible et l'optimisation de celui-ci au moyen d'un système de règles de transformation. L'étude et les développements s'appuient sur l'extension de la bibliothèque java appelée XXL (http://www.xxl-library.de/). Des notions de gradualité (ensembles flous, prédicats flous, opérateurs algébriques étendus…) y sont introduites afin de pouvoir traiter des requêtes flexibles. XXL est par ailleurs utilisée dans le cadre de la mise en œuvre du système PreferenceSQL [14]. Les objectifs de ce système se démarquent de notre approche par le fait que les préférences considérées ne sont pas toujours commensurables (composables). Ainsi, XXL offre une plateforme intéressante pour développer, expérimenter et comparer différents algorithmes d'évaluation.

Remerciements

Ce travail a été partiellement supporté par le Ministère de la Recherche et des Nouvelles Technologies, programme ACI Masses de

Données, projet MD-33, Accès Personnalisé à des Masses des Données (APMD).

Références

[1] Albert J. , « Algebraic properties of bag data types », *17th Inter. Conf. On Very Large Data (VLDB'91)*, Barcelona, p. 211-219, 1991.

[2] Blizard W. D., *Generalizations of the concept of sets, a formal theory of multisets*, PhD thesis, University of Oxford, 1985.

[3] Börzsönyi S., Kossmann D., Stocker K., The Skyline operator, *17th Int. Conf. On Data Engineering*, p. 421-430, 2001.

[4] Bosc, P., Liétard L., A general technique to measure gradual properties of fuzzy sets, *International Conference of the Fuzzy Sets Association (IFSA'05)*, 2005.

[5] Bosc P., Liétard, L., Pivert O. Quantified statements and database fuzzy querying. In P. Bosc and J. Kacprzyk (eds), *Fuzziness in Database Management Systems*: (p. 275-308). Physica-Verlag, 1995.

[6] Bosc P., Pivert O., SQLF : a relational database language for fuzzy querying , *IEEE Transactions on Fuzzy Systems*, vol. 3, p. 1-17, 1995.

[7] Chomicki J., Preference formulas in relational queries, *ACM Trans. On Database Systems*, 27, p. 153-187, 2003.

[8] Connan F., *Interrogation flexible de bases de données multimédia*, Thèse de doctorat, Université de Rennes I, 1999.

[9] Dubois D., Prade H., Fuzzy cardinality and modeling of imprecise quantification , *Fuzzy Sets and Systems* , vol. 16, n°3, p. 199-230, 1985.

[10] Dubois D., Prade H., Fuzzy numbers: an overview, *Analysis of fuzzy information, Mathematics and Logics*, vol.I, p. 3-39, 1987.

[11] Glökner I., *Fuzzy quantifiers in natural language : semantics and computationnal models*. Der Andere Verlag, 2004.

[12] Ichikawa T., Hirakawa M., ARES: a relational database with the capability of performing flexible interpretation of queries, *IEEE Transactions on Software Engineering*, 2, p. 624-634, 1986.

[13] Kacprzyk J., A. Ziolkowski, Database queries with fuzzy linguistic quantifiers, *IEEE Transactions on Systems, Man and Cybernetics*, 16, p. 474-478, 1986.

[14] Kieβling W. Köstler G., PreferenceSQL – Design, implementation, experiences., *28th Inter. Conf. On Very Large Data Bases*, p. 990-1001, 2002.

[15] Klement E. P., Mesiar R., Pap E., Triangular norms. Position paper II: general constructions and parameterized families, *Fuzzy Sets and Systems*, vol. 145, p. 411-438, 2004.

[16] Knuth D. E., *The art of computer programming*, vol. 2, Addison Wesley, 1985.

[17] Lacroix M., Lavency P., 1987. Preferences: putting more knowledge into queries, *13th Very Large Data Bases Conference*, p. 217-225.

[18] Lamperti G. Melchiori M., Zanella M., « On multisets in database systems », *Multisets processing, Lecture Notes in Computer Science LNCS 2235*, Springer-Verlag, 2001, p. 147-215.

[19] Liétard L., Rocacher D., A generalization of the OWA operator to evaluate non monotonic quantifiers, *4th Conference of the European Society for Fuzzy Logic and Technology and 11th Rencontres Francophones sur la Logique Floue et ses Applications*, p. 183-188, 2005.

[20] Liu Y., Kerre E., An overview of fuzzy quantifiers, part 1 : interpretations. *Fuzzy Sets and Systems*, 95, Issue 1, p. 1-21, 1998.

[21] Miyamoto S., Fuzzy Multisets and Fuzzy Clustering of Documents, *10th Inter. Conf. On Fuzzy Systems* FUZZ IEEE'01, 2001.

[22] Motro A., VAGUE: a user interface to relational databases that permits vague queries, ACM Trans. On Office Information Systems, 6, p. 187-214, 1988.

[23] Pivert O., *Contribution à l'interrogation flexible de bases de données : expression et évaluation de requêtes floues*, thèse de doctorat, Université de Rennes 1, 1991.

[24] Rocacher D., *Multi-ensembles flous et quantités graduelles – Application à l'interrogation flexible de bases de données*, Habitation à Diriger des Recherches, Université de Rennes 1, décembre 2005.

[25] Rocacher D., On fuzzy bags and their application to flexible querying, *Fuzzy Sets and Systems* 140 (1), p. 93-110, 2003.

[26] Rocacher D., Bosc P., About $\mathbb{Z}_f$, the set of fuzzy relative integers, and the definition of fuzzy bags on $\mathbb{Z}_f$, *Lecture Notes in Computer Science* LNCS 2715, Springer-Verlag, p. 95-102, 2003.

[27] Rocacher D., Bosc P., The set of fuzzy rational numbers and flexible querying, *Fuzzy Sets and Systems*, 155, p. 317-339, 2005.

[28] Rocacher D., Bosc P., The set of fuzzy relative integers and fuzzy bags, *International Journal of Intelligent System IJIS,* à paraître.

[29] Rocacher D., Connan F., Flexible queries in object databases: on the study of bags, *8th Inter.l Conf. on Fuzzy Systems*, p. 615-620, 1999.

[30] Rocacher D., Connan F., Les multi-ensembles flous un concept générique pour l'interrogation flexible de bases de données objet, *16e Journées Bases de Données Avancées*, p. 83-96, 2000.

[31] Schweizer B., Sklar A., Associative functions and abstract semi-groups, *Publicationes Mathematicae Debrecen*, 10, p. 69-81, 1963.

[32] Whalen T., Parameterized R-implications, *Fuzzy Sets and Systems*, vol. 134, p. 231-281, 2003.

[33] Wygralak M., Questions of cardinality of finite fuzzy sets, *Fuzzy Sets and Systems*, vol. 102, p. 185-210, 1999.

[34] Yager R., « On the theory of bags », *Inter. Journal of General Systems*, vol. 13, p. 23-27, 1986.

[35] Yager, R.R., On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *Transac. on Syst., Man, and Cybernetics*, 18, p. 183-190, 1988.

[36] Zadeh L. A., A computational approach to fuzzy quantifiers in natural languages, *Comp. Math. App.*, vol. 9, p. 149-184, 1983.