APPROXIMATION LINGUISTIQUE APRES RAISONNEMENT APPROXIMATIF

G. ZUNINO*, A.M. DESODT, D.JOLLY
Centre d'Automatique de Lille
Université des Sciences et Technologies de Lille
59655 Villeneuve d'Ascq Cedex

* HEI
13, rue de Toul
59046 LILLE CEDEX

FRANCE

Abstract

After using a approximate raisonning method (like PRUF [ZAD. 1978] for instance), we sometimes need a linguistic approximation of the resulting fuzzy set. We propose two alternate methods of the P. BONISSONE process. First, we will use distances between two fuzzy sets and second, we will use the divergence measure. Those ameliorations simplify the approximation process and sometimes improve the calculation speed. An example of application will be described.

Mots Clés : approximation linguistique, distances entre sous-ensembles flous, divergence, PRUF.

Keywords : linguistic approximation, distances between fuzzy sets, divergence, PRUF.

1. Introduction

Actuellement, l'automatisation d'un système tend à réintroduire l'homme dans le processus. Par exemple, l'opérateur peut informer la machine en utilisant un langage naturel comme le français. L'interface utilisateur devra donc convertir une phrase du langage naturel en une représentation utilisable par la machine. Le langage naturel présentant des imprécisions et des incertitudes, il est possible de le représenter par un sous-ensemble flou. Inversement, lorsque la machine désirera informer l'opérateur après un processus de prise de décision, il faudra convertir les sous-ensembles flous en une phrase du langage naturel. Ce document présente deux améliorations de la méthode de P. BONISSONE [BON. 1979] destinée à déterminer l'approximation linguistique d'un sous-ensemble flou.

2. Le problème

Le processus d'approximation linguistique consiste à interpréter le sens d'une fonction d'appartenance et à lui associer un label appartenant à un ensemble de mots V. P. BONISSONE présente une méthode dont les étapes sont brièvement rappelées :

- Définir l'ensemble V des mots. Cet ensemble sera appelé "vocabulaire".
- Associer une fonction d'appartenance à chaque mot de V. Ces fonctions d'appartenance seront dites "de référence".
- Se donner une fonction F qui à un sous-ensemble flou S, associe un vecteur de l'espace $\Re$ de dimension k. Nous supposerons que k est petit (par exemple k = 4).
- Calculer l'image par F de toutes les fonctions de référence.
- Calculer l'image par F du sous-ensemble inconnu dont on désire une approximation.
- Fixer un seuil ε .

- Créer l'ensemble AL des fonctions d'appartenance de référence telles que la distance Euclidienne pondérée entre le point représentant le sous-ensemble inconnu et tous les points représentant les fonctions de référence soit inférieure au seuil.
- Fixer un second seuil ε'.
- Parmi les éléments de AL, regrouper dans un ensemble AL' ceux dont la distance d (à choisir) avec le sous-ensemble flou inconnu est inférieure à ε'.

L'ensemble AL' contient donc zéro, un ou plusieurs mots de V dont les significations sont proches de celle de l'ensemble flou inconnu. La précision de l'approximation dépend évidemment du choix de ε et de ε'. Si ε est trop grand, AL risque de contenir presque tous les mots du vocabulaire. Dans le cas contraire, si ε est trop faible AL sera vide. Il en est de même pour le choix de ε' relativement à l'ensemble AL'. Le choix de ces seuils doit être fait de façon pragmatique et résulte souvent de nombreux essais.

Cette méthode nécessite donc la détermination d'une fonction F qui doit représenter un sous-ensemble flou dans $\Re^k$ avec le plus de précision possible. ce choix est souvent délicat. Il faut aussi choisir une distance d et se fixer deux seuils d'approximation.

Par ailleurs, si F n'est pas injective, deux sous-ensembles flous peuvent être représentés par le même point de $\Re^k$. Pourquoi ne pas évaluer la similarité sémantique de deux sous-ensembles flous directement, sans passer dans un ensemble intermédiaire ? Ceci supprime le choix de F , le problème lié à la non injectivité et le choix d'un seuil d'approximation.

Pour évaluer la similarité entre deux sous-ensembles flous, il existe au moins deux méthodes. Il est possible d'utiliser une distance entre sous-ensembles flous ou une mesure de divergence [DIN. 1993].

3. Utilisation d'une distance

Soit V un ensemble de mots. Chacun de ces mots est associé à une fonction d'appartenance de référence. L'algorithme d'approximation proposé après le choix d'une distance d est le suivant :

- Fixer un seuil ε
- Former l'ensemble AL des mots dont la distance d entre la fonction de référence et le sous-ensemble inconnu est inférieure à ε.
- AL est le résultat cherché

Dans ce cas, il se pose un problème de choix de distance entre sous-ensembles flous. Toutes les distances sont potentiellement candidates. Cependant, la distance choisie doit évaluer correctement la similarité sémantique entre deux ensembles. Ici, l'étude se limitera à quelques distances classiques. A et B sont deux sous-ensembles flous de fonctions d'appartenances respectives μ_A et μ_B. x_i sont des éléments de A et B.

- Distance de Hamming

$$d_{Hm}(A,B) = \sum_{i=1}^{n} \left| \mu_A(x_i) - \mu_B(x_i) \right|$$

- Distance Euclidienne

$$d_{Eu}(A,B) = \sqrt{\sum_{i=1}^{n} (\mu_A(x_i) - \mu_B(x_i))^2}$$

- Distance Euclidienne pondérée

$$d_{Ep}(A,B) = \sqrt{\sum_{i=1}^{n} W_i^2 . (\mu_A(x_i) - \mu_B(x_i))^2}$$

W_i étant un réel

• Distance l^p de Kosko (p_1)

$$l^p(A,B) = \sqrt[p]{\sum_{i=1}^{n} |\mu_A(x_i) - \mu_B(x_i)|^p}$$

• Distance de Bhattacharyya

Soient

$$p_A(x_i) = \frac{\mu_A(x_i)}{\sum \mu_A(x_i)} \text{ et } p_B(x_i) = \frac{\mu_B(x_i)}{\sum \mu_B(x_i)}$$

Définissons alors le coefficient de Bhattacharyya R comme

$$R(p_A, p_B) = \sum_{i=1}^{\text{Card } U} \sqrt{p_A(x_i).\, p_B(x_i)}$$

R varie entre 0 et 1. L'égalité avec l'unité est atteinte lorsque A et B sont identiques. La distance de Bhattacharyya est alors définie par

$$d_{Bat}(A,B) = \sqrt{1 - R(p_A, p_B)}$$

Nous pouvons remarquer que $0 \leq d_{Bat}(A,B) \leq 1$.

La distance de Hamming permet d'évaluer un index de similarité plutôt qu'une mesure de la proximité de deux ensembles, elle ne sera donc pas utilisée en approximation linguistique.

La distance Euclidienne mesure bien la ressemblance entre deux sous-ensembles flous. Elle sera donc utilisable dans notre approximation linguistique.

La distance Euclidienne Pondérée est comparable à la distance Euclidienne mais permet l'introduction de coefficients W_i pour donner plus d'importance à la différence $\mu_A(x_i) - \mu_B(x_i)$ pour un x_i donné. Les W_i permettent aussi de normaliser le résultat de cette distance entre 0 et 1.

Le choix des coefficients W_i dépend de l'application et ils sont souvent fixés par plusieurs essais. Cette distance semble donc - en général - difficile à utiliser et ne sera pas considérée par la suite.

La distance l^p de Kosko est une généralisation de la distance Euclidienne. En général, p est un réel choisi dans l'intervalle [1,4]. Cette distance présente peu d'intérêt pour l'approximation linguistique. En effet, l'élévation à une puissance réelle peut conduire à des approximations numériques nuisibles.

La distance de Bhattacharyya est issue de la détection de signaux [KAIL. 1967, CHER. 1952]. Elle permet d'évaluer la similarité sémantique entre deux ensembles flous et par conséquent, il s'agit de la distance la mieux adaptée pour l'approximation linguistique.

Contrairement à l'algorithme de P. BONISSONE, il **n'y a pas** ici de "premier tri". La distance joue donc un rôle déterminant dans l'approximation. En effet, si la distance mesure plutôt la proximité des ensembles, il est possible que AL contienne des éléments dont la signification n'est pas proche de celle du sous-ensemble inconnu.

3.1. Distance Euclidienne

Elle sera utilisée lorsque les fonctions de références seront bien séparées les une des autres.

Deux fonctions d'appartenance seront dites "**séparées d'un seuil** λ" lorsque la distance Euclidienne entre les deux sous-ensembles flous correspondants sera supérieure à λ.

Pour que cette distance soit utilisable, il faudra donc un **vocabulaire simple aux mots bien séparés**. Le nombre de mots n'a aucune influence sur cette méthode. La distance Euclidienne sera adaptée à des applications où la vitesse d'approximation est primordiale et où la précision peut être moindre.

3.2. Distance de Bhattacharyya

Cette distance évalue très bien la similarité sémantique entre deux sous-ensembles flous. Elle sera donc utilisée lorsque les limites de la distance Euclidienne auront été atteintes. Par exemple pour des vocabulaires où les fonctions de références sont mal séparées. Cependant, cette distance nécessite un temps de calcul plus long que la distance Euclidienne.

La distance de Bhattacharyya sera utilisée pour des applications où il est nécessaire d'avoir un bon compromis Vitesse/Précision.

4. Utilisation de la divergence

La divergence entre deux ensembles flous A et B mesure la facilité avec laquelle ils peuvent être discernés. Voici son expression [DIN. 1993] :

$$D(A,B)=\sum_{i=1}^{n}\left[(\mu_A(x_i)-\mu_B(x_i)).\ln\frac{1+\mu_A(x_i)}{1+\mu_B(x_i)}+(\mu_B(x_i)-\mu_A(x_i)).\ln\frac{2+\mu_A(x_i)}{2+\mu_B(x_i)}\right]$$

Cette divergence possède des propriétés très intéressantes :

- $D(A,B) \geq 0$,
- $D(A,B)=0 \Leftrightarrow A=B$
- $D(A,B) = D(B,A)$
- $D(A \cup B, A \cap B) = D(A,B)$
- $\forall A,B,C \in U,\ D(A \cup B, C) \leq D(A,C)+D(B,C)$
- Posons A = C. A l'aide de la propriété précédente, on obtient : $D(A,B) \geq D(A \cup B, A)$

Les trois dernières propriétés sont très utiles en vision artificielle et en reconnaissance des formes. Elles sont aussi utiles en approximation linguistique.

L'algorithme d'approximation linguistique est le suivant :

- Définir l'ensemble V des mots ainsi que les fonctions de référence.
- Fixer un seuil ε
- Créer l'ensemble AL des fonctions d'appartenance de référence telles que la divergence entre le point représentant le sous-ensemble inconnu et tous les points représentant les fonctions de réference soit inférieure au seuil.
- Si l'approximation est jugée satisfaisante

Arrêt : AL est le résultat

Sinon

  - Fixer un second seuil ε'
  - Parmi les éléments de AL, regrouper dans un ensemble AL' ceux dont la distance d (à choisir) avec le sous-ensemble flou inconnu est inférieure à ε'.
  - AL' est le résultat

Cet algorithme est plus proche de celui de P. BONISSONE que le précédent. En effet, il y a d'abord **présélection** des mots puis approximation linguistique sur les éléments présélectionnés. Le tri préalable permet de ne garder dans AL que les mots dont le sens est proche du sous-ensemble inconnu. Parfois, pour des applications simples, ce tri préalable est suffisant pour obtenir l'approximation. Dans les cas plus complexes, l'utilisation d'une distance fournira le résultat cherché.

Cette alternative permet de se rapprocher du processus de P. BONISSONE sans avoir le problème de la détermination de la fonction F. Le calcul de la divergence est souvent plus rapide à réaliser que celui induit par la fonction F.

Si il faut utiliser une distance, les distances candidates sont celles décrites au paragraphe précédent. En pratique, les distances Euclidienne ou distance de Bhattacharyya seront utilisées.

Cette alternative est réservée aux cas où la précision d'approximation est importante vis à vis de la vitesse.

5. Exemple

La téléopération consiste à conduire un bras esclave dans un milieu hostile à l'homme à l'aide d'un bras maître de manière à réaliser une tache déterminée. Il est également possible de conduire le bras esclave d'une manière semi-automatique. Dans ce cas, l'opérateur exécute une partie de la tâche en manuel alors que l'autre est réalisée entièrement sous le contrôle de la machine. Dans ce contexte, il faut à chaque instant déterminer quel mode (manuel, semi-automatique ou automatique) est le plus adapté. Pour prendre cette décision, la machine sera renseignée sur le milieu extérieur par les caractéristiques suivantes [WAW. 1993 A] :

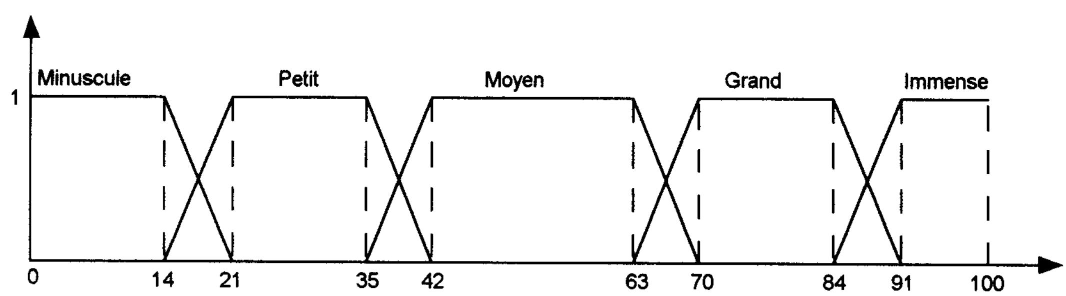
Critère Principal	Opérateur	Télémanipulateur	Information	Tâche
Premier sous-critère	Performance de l'opérateur	Confiance	Quantité	Simplicité
Deuxième sous-critère	Vigilance	Performance du Télémanipulateur	Qualité	Temps

Chacun de ces critères sera mesuré en temps réel et comparé avec les préférences de l'expert [WAW. 1993 B].

Chacun de ces critères sera modélisé par un vocabulaire. Par la suite, nous convertirons une phrase d'un langage naturel en un sous-ensemble flou et nous en chercherons l'approximation linguistique puis nous comparerons les trois méthodes.

5.1. Le vocabulaire

Pour faciliter l'approximation linguistique, il est judicieux de choisir le même vocabulaire pour chacun des critères. Par ailleurs, les fonctions d'appartenance de chacun des mots du vocabulaire seront choisies trapézoïdales. Le vocabulaire sera constitué de 5 adjectifs de base :



Ce vocabulaire sera complété par 4 modificateurs :

- Très
- Plutôt

- Plus ou moins
- Pas

Chacun des adjectifs de base sera combiné avec chaque modificateur. Notre vocabulaire comportera donc 25 mots.

Chaque modificateur agit sur la fonction d'appartenance du mot qui le suit en accord avec la méthode PRUF [ZAD. 1978]. Soit f la fonction d'appartenance du mot suivant le modificateur :

• Très	f^2
• Plutôt	$f^{1.75}$
• Plus ou moins	$f^{0.5}$
• Pas	$1-f$

Par exemple, Très Moyen aura la fonction d'appartenance de Moyen élevé au carré.

Pour les résultats présentés, nous avons calculé l'approximation linguistique du Mot "Très Très Grand". Le sous-ensemble flou de "Très Très Moyen" a été obtenu à l'aide de la méthode PRUF.

5.2. Utilisation d'une distance

Voici les résultats par l'alternative utilisant une distance :

Distance Euclidienne

Mot du vocabulaire	Distance avec le se flou inconnu
Très Grand	0.595487108913
Plutôt Grand	0.719293372970
Grand	1.246528641829
Plus-ou-moins Grand	1.854395671420
Non Moyen	2.500708516295
Non Immense	2.500708516295
Non Petit	3.309569837916
Non Minuscule	3.309569837916
Plutôt Petit	3.967787357211
Plus-ou-moins Petit	3.967787357211
.	.
.	.

Nous remarquons qu'une approximation de "Très Très Grand" est "Très Grand". Si nous fixons le seuil à 1.90, il y a quatre mots de notre vocabulaire qui approximent le sous-ensemble flou inconnu.

Cependant, la distance Euclidienne induit des groupements néfastes. Par exemple, il est impossible de savoir si "Plus ou Moins Petit" est une meilleure approximation que "Plutôt Petit".

Distance de Bhattacharyya

Mot du vocabulaire	Distance avec le se flou inconnu
Très Grand	0.108671120398
Plutôt Grand	0.128603528944
Grand	0.203023742174
Plus-ou-moins Grand	0.268942831596
Non Moyen	0.324555065238
Non Immense	0.324555065238
Non Petit	0.393113278901
Non Minuscule	0.393113278901
Non Grand	0.910789270482

Plus-ou-moins Immense	0.924773740874
Plus-ou-moins Moyen	0.924773740874
Moyen	0.937797698916
Immense	0.937797698916
Plutôt Immense	0.953693753951
Plutôt Moyen	0.953693753951
Très Immense	0.957995379820
Très Moyen	0.957995379820
Plus-ou-moins Minuscule	1.000000000000
Plus-ou-moins Petit	1.000000000000
Minuscule	1.000000000000
Petit	1.000000000000
Très Minuscule	1.000000000000
Plutôt Petit	1.000000000000
Plutôt Minuscule	1.000000000000
Très Petit	1.000000000000

Distance la plus petite = 0.108671120398 pour le mot "Très Grand"

La distance de Bhattacharyya donne là encore un résultat correct. Les distances sont comprises entre 0 et 1 ce qui peut poser un problème lorsque le sous-ensemble flou inconnu est très différent de tous les mots du vocabulaire. Le phénomène de groupement est existant mais moins marqué en général.

5.3. Utilisation de la divergence

Voici les résultats après l'évaluation de la divergence. Dans ce cas, il n'a pas été nécessaire de réaliser la deuxième étape de l'algorithme :

Divergence

Mot du vocabulaire	Distance avec le se flou inconnu
Très Grand	0.486363194848
Plutôt Grand	0.709445792226
Grand	2.128389157492
Plus-ou-moins Grand	4.704938050278
Non Moyen	8.638533060342
Non Immense	8.638533060342
Non Petit	15.148676963192
Non Minuscule	15.148676963192
Plutôt Petit	21.829816678856
Plus-ou-moins Petit	21.829816678856
Plus-ou-moins Minuscule	21.829816678856
Minuscule	21.829816678856
Petit	21.829816678856
Très Petit	21.829816678856
Très Minuscule	21.829816678856
Plutôt Minuscule	21.829816678856
Très Immense	24.334574856801
Très Moyen	24.334574856801
Plutôt Immense	24.479803442458
Plutôt Moyen	24.479803442458
Immense	25.148476205794
Moyen	25.148476205794
Plus-ou-moins Immense	26.102193933604

Plus-ou-moins Moyen	26.102193933604
Non Grand	28.467135732733
Divergence la plus petite = 0.486363194848 pour le mot "Très Grand"	

La divergence donne aussi une approximation correcte. Les groupements sont plus marqués que pour les deux méthodes précédentes. Ceci montre bien que la divergence réalise avant tout un premier tri.

5.4. Comparaison des alternatives

Les méthodes présentées sont adaptées pour obtenir l'approximation après un raisonnement utilisant PRUF. Le vocabulaire utilisé pour la méthode PRUF et celui utilisé par l'approximation linguistique **doivent** être les même. Si tel n'est pas le cas, l'approximation linguistique fournira des résultats aberrants.

Ces alternatives consomment en général moins de temps de calcul que la méthode originale. La méthode utilisant une distance peut s'exécuter jusqu'à 50% plus vite que la méthode de P. BONISSONE tandis que la méthode de la divergence est environ plus rapide de 20% en moyenne.

En ce qui concerne la robustesse de ces méthodes, une petite perturbation d'une fonction de référence ne perturbe pas du tout les méthodes. Il en est de même pour une petite perturbation du sous-ensemble flou inconnu.

Pour un sous-ensemble continu, le nombre de points d'appartenance fournis peut augmenter sans faire croître la complexité de l'algorithme. Il est ainsi possible d'augmenter la précision de l'ensemble inconnu sans augmenter beaucoup le temps nécessaire au calcul.

De plus, nous avons travaillé sur des sous-ensembles discrets, mais il est possible d'envisager d'utiliser ces techniques pour des ensembles continus.

Ces méthodes présentent un désavantage. Il faut analyser au moins une fois tout le vocabulaire pour chaque sous-ensemble inconnu. Lorsque le vocabulaire grandit, le temps d'exécution peut dépasser les limites du raisonnable.

6. Conclusion

Nous avons défini deux alternatives à la méthode d'approximation linguistique de P. BONISSONE. Celles-ci simplifient l'algorithme orignal, réduisent l'espace mémoire nécessaire et diminuent le temps de calcul. Les méthodes présentées sont dépendantes de distances entre sous-ensembles flous. Il sera donc nécessaire de déterminer la distance la plus adaptée à une application particulière. Ces nouveaux procédés ont été mis en oeuvre avec succès dans le cadre d'une interface Homme/Machine d'un système manufacturier où le temps d'exécution est primordial. D'autres travaux en cours s'intéressent à la recherche de distances spécialement adaptées à la reconstitution linguistique.

Bibliographie

[ZAD. 1978] Zadeh L.A. (1978). PRUF - A meaning representation language for natural languages. *Int. J. Man-Machine Studies, 395-160.*

[BON. 1979] Bonissone P. (1979). A pattern recognition approach to the problem of linguistic approximation in system analysis. *Computer Science Divison. Department of Electrical Ingineering and Computer Science. University of California, Berkeley.*

[DIN. 1993] Dinabandhu Bhandari, Nikhil R.Pal. (1993) Some new Information Measures for Fuzzy Sets. *Information Sciences 67,209-228.*

[KAIL. 1967] Kailath T. (1967). The divergence and Bhattacharyya distance measure in signal détection. *IEEE Trans. Commun. Tech. 15, No.1, pp. 609-637.*

[CHER. 1952] Chernoff H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Ann. Math. Stat., vol 23, pp.493-507.*

[WAW. 1993 A] Wawak F., Desodt A.M., Jolly D. (1993). Fuzzy decision algorithm for man-machine systems. *Qualitive Raisonning and decision Technologies. Barcelona 1993.*

[WAW. 1993 B] Wawak F., Desodt A.M., Jolly D. (1993). Des opérateurs d'agrégation floue "horizontaux" pour l'aide à la décision en téléopération. *BUSEFAL 54 - Printemps 1993.*