Apprentissage flou

Ref : * Michel Grabisch, 2020
* Louis Gacogne, "Éléments de logique floue"
* Fuzzy models for Pattern Recognition

I introduction

La théorie des ensembles flous est développée en 1965 par Lotfi Zadeh. Il s'agit d'une théorie du domaine de l'algèbre abstraite qui a pour but de représenter mathématiquement l'imprécision relative à certaines classes d'objets. Cette théorie sert de fondement à la logique floue.

Très rapidement, on s'est aperçu que la notion de classes utilisée en apprentissage (notamment pour la reconnaissance de formes) trouvait là un cadre naturel. En effet une classe est souvent représentée comme un groupe d'individus ayant des similitudes. Ces similitudes peuvent être plus ou moins fortes entre les individus d'une même classe, et un individu peut peut présenter des similitudes avec des individus d'autres classes. Ainsi, l'appartenance d'un individu peut ne pas être réduite à une unique classe mais "distribuée" sur plusieurs classe. Cela rejoint le formalisme des ensembles flou puisque* un élément peut appartenir plus ou moins fortement à plusieurs ensembles flous. Un exemple simple en classification botanique est le "nectarine" qui est un croisement entre la prune et la pêche.

Cependant, si les ensembles flous peuvent sembler constituer un cadre naturel pour la ~~cla~~ notion de classe, force est de constater que les algorithmes classiques considèrent implicitement les classes commes "nettes". Les Méthodes Bayesiennes donnent bien une probabilité d'appartenir à une classe mais cette notion de probabilité (associée à la fréquence d'apparition d'un événement lorsqu'on répète une expérience) est fondamentalement différente de la notion d'appartenance graduée à un ensemble (notion ~~sur~~ laquelle repose ~~Plus~~ la théorie "flou"). Un exemple simple (cf "Fuzzy models for Pattern recognition") illustre bien les choses:

Exemple

Considérons $\mathcal{L}$ l'ensemble des liquides, $\mathcal{L}$ le sous ensemble flou des liquides potable. On est dans le désert, sans avoir bu depuis 1 semaine. On trouve deux bouteilles A et B et il faut choisir laquelle boire. On a les informations suivantes:

* le degré d'appartenance de B à $\mathcal{L}$ est 0.9 (échelle [0,1])
* la proba que A appartienne à $\mathcal{L}$ est 0.9

Si on choisi B, on sait que le liquide est "pratiquement potable". Ce qui veut dire qu'il ne contiendra pas les substances les + toxiques

Si on choisi A, on a 9 chances sur 10 d'avoir 1 liquide parfaitement potable, mais 1 chance sur 10 d'avoir un poison

Nous allons commencer par présenter les notions de base de la théorie des ensembles flous et de la logique flou. Dans un second temps nous aborderons les techniques de classification floues

II Ensembles flous - Relation floues

II.1 Ensembles flous (1965)

Dans le cadre de la théorie des ensembles conventionelle, les éléments ~~appartiennent~~ ~~sont ou~~ appartiennent ou non (de manière binaire) à des ensembles. Dans le cadre de la théorie floue, l'appartenance d'un élément à un ensemble n'est pas si tranchée. Pour représenter ceci mathématiquement, on utilisant le degré d'appartenance d'un élément à un ensemble, les ensembles flous sont définis de la façon suivante :

Def III.1 : Ensembles flous

Un ensemble flou ~~A~~ est un ensemble charactérisé par une fonction d'appartenance $f_A : X \longrightarrow [0, 1]$, où X est un espace quelconque

Si $f_A(x)$ est proche 1, x a un fort degré d'appartenance à ~~A~~

Si $f_A(x)$ ——— 0, x ——— faible ——— A

Si $f_A : X \longrightarrow \{0, 1\}$ on retombe sur la théorie classique

Def II.2 : Ensemble vide

A (ensemble flou) est vide ssi $f_A \equiv 0$

Def II.3 : Egalité

2 ensembles flous A et B sont égaux ssi $f_A \equiv f_B$ $(\Leftrightarrow \forall x, f_A(x) = f_B(x))$

Def II.4 : Complémentaire

A un ~~Ensemble~~ ensemble flou, son complémentaire A' est défini par $f_{A'}(x) = 1 - f_A(x), \forall x \in X$

Def II.5 Inclusion

A, B ensembles flous. A est inclu dans B $(A \subset B)$ ssi $f_A(x) \leq f_B(x)\ \forall x \in X$

Def II.6 : Union

A, B ~~deux~~ ensembles flous. $C = A \cup B$ ~~est a fu~~ est un ensemble flou défini par $f_C(x) = \max\{f_A(x), f_B(x)\}, \forall x \in X$

Proposition II.1

Comme dans le cadre classique :

$A \cup (B \cup C) = (A \cup B) \cup C$

$A \subset A \cup B$

$B \subset A \cup B$

Preuve : exercice

Def II.7 : Intersection

L'intersection de deux ensembles flous A et B est un ensemble flou $C = A \cap B$ défini par $f_C(x) = \min\{f_A(x), f_B(x)\}$ $\forall x \in X$

Proposition II.2

Comme dans le cadre classique :

$A \cap (B \cap C) = (A \cap B) \cap C$

$A \supset A \cap B$

$B \supset A \cap B$

Preuve : Exercice

Proposition II.3

$(A \cup B)' = A' \cap B'$

$(A \cap B)' = A' \cup B'$

$C \cap (A \cup B) = (C \cap A) \cup (C \cap B)$

$C \cup (A \cap B) = (C \cup A) \cap (C \cup B)$

Preuve : Exercice

II.2 : Relations floues

$D \subset \mathbb{R}^d$, $k > 1$. Commençons par le cadre conventionnel

une partition de D en k cluster/classes est donnée par des fonctions indicatrices $\gamma_1, \ldots, \gamma_k$:

$$\gamma_i(x) = \begin{cases} 1 & \text{si } x \in \text{classe } i \\ 0 & \text{sinon} \end{cases} \qquad i \in \{1, \ldots, k\}$$

une relation "dure" r dans D est définie comme une fonction $r : D \times D \to \{0, 1\}$, et $x, y \in D$ sont dits partageant une relation si $r(x,y) = 1$

formalisation matricielle : $D = \{x_1, \ldots, x_n\} \in (\mathbb{R}^d)^n$

$$\gamma_i(x_j) \equiv \gamma_{i,j} \quad \begin{pmatrix} i \in \{1, \ldots, k\} \\ j \in \{1, \ldots, n\} \end{pmatrix} ; \; r(x_i, x_j) = r_{ij} \quad \begin{pmatrix} i \in \{1, \ldots, n\} \\ j \in \{1, \ldots, n\} \end{pmatrix}$$

$\mathcal{M}_k$ = ensemble de toutes les matrices de partition U

$$\begin{cases} \dim(U) = k \times n \\ U[i,j] = \gamma_{i,j} \in \{0,1\} \\ \sum_{i=1}^{k} \gamma_{i,j} = 1 \quad \forall j \quad \text{(Chaque donné appartient à une unique ~~cluster~~ classe} \\ \sum_{j=1}^{n} \gamma_{i,j} > 0 \quad \forall i \quad \text{(pas de classe vide)} \end{cases}$$

Pour $U \in \mathcal{M}_k$, il y a une matrice de relations $R = (r_{j,\ell})$ avec

$\dim(R) = n \times n$, définie par

$$r_{j,\ell} = \begin{cases} 1 \text{ si } \gamma_{i,j} = \gamma_{i,\ell} = 1 \text{ pour 1 certain } i \text{ (i.e. } x_j, x_\ell \text{ sont dans une m̂ classe } i) \\ 0 \text{ sinon} \end{cases}$$

* Cadre flou :

Def II.8 : Matrice de partition floue

$U \in \mathcal{M}_{f,k}$ ($\dim(U) = k \times n$, $\mathcal{M}_{f,k}$ = ensemble de toutes les matrices de partition floue)

$\boxed{U[i,j] = \gamma_{i,j} \in [0,1]}$

$\sum_{i=1}^{k} \gamma_{i,j} = 1 \ \forall j$

$\sum_{j=1}^{n} \gamma_{i,j} > 0 \ \forall i$

Il est possible de définir une matrice de relations floues R_f associé à U. par

Def II.9 : Matrice de relations floues

$U \in \mathcal{M}_{f,k}$. La matrice de relations floue associée R_f est définie par

$$R_f[j,\ell] = r_{j,\ell} = \max_i \left(\min(\gamma_{i,j}, \gamma_{i,\ell})\right)$$

$\dim(R_f) = n \times n$

Exemple :

* 3 données, 2 clusters/classes

$$U_f : \begin{pmatrix} 0,3 & 0,2 & 0,9 \\ 0,7 & 0,8 & 0,1 \end{pmatrix} \leadsto$$

x_1 plutôt dans la classe 2
x_2 plutôt ——— 2
x_3 ——— 1

matrice de relation équivalente

$$R_f = \begin{pmatrix} 0,7^{(*)} & 0,7 & 0,3 \\ 0,7 & 0,8 & 0,2 \\ 0,3 & 0,2 & 0,9 \end{pmatrix}$$

$(*) = \max(\min(\mu_{1,1}, \mu_{1,1}), \min(\mu_{21}, \mu_{21}))$

si on oublie la diagonale,
x_1 et x_2 sont plutôt ensembles
$\left.\begin{array}{l}(x_2, x_3)\\(x_1, x_3)\end{array}\right\}$ plutôt pas ensembles

~~On va maintenant essayer d'adapter les algorithmes~~

~~III Classification floue~~

→ équivalent classique (ou "dur")

$$U = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \leadsto R = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

III Classification floue :

III.1 kppv

$S = \{x_1, \ldots, x_n\}$ dont les labels sont connus

→ $\mu_{ij} \in \{0,1\}$ (comme dans le cadre classique)

→ $r_{j,l} \in \{0,1\}$

$S' = \{x'_1, \ldots, x'_{n'}\}$ dont les labels sont inconnus

On va recalculer les $\mu'_{i,t}$ $\quad i \in \{1, \ldots, \text{nb de classes}\}$, $t \in \{1, \ldots, n'\}$

$$\mu'_{i,t} = \frac{\sum_{j=1}^{k} \mu_{i,j} \times \frac{1}{\|x'_t - x_j\|^{2/(\partial-1)}}}{\sum_{j=1}^{k} \frac{1}{\|x'_t - x_j\|^{2/(\partial-1)}}}$$

(k = nbre de Voisins)

(+ facile de commencer par $\mu_i(x'_t)$ → nouvelle donnée
↳ n° de la classe

∂ est un paramètre ($1<\partial<\infty$) qui détermine un degré de flou de la classification. (On le testera avec R)

Exemple 3 données, 2 classes — 2 ppv

$$U = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \qquad \begin{aligned} x_1 &= (0,0) \\ x_2 &= (1,1) \\ x_3 &= (4,4) \end{aligned}$$

deux nouvelles données : $x'_1 = (0, 1)$
$x'_2 = (4,1)$

les 2 ppv de x'_1 : x_1 et x_2 $\begin{cases} \|x'_1 - x_1\| = 1 \\ \|x'_1 - x_2\| = 1 \end{cases}$

les 2 ppv de x'_2 : x_2 et x_3 $\begin{cases} \|x'_2 - x_2\| = 3 \\ \|x'_2 - x_3\| = 3 \end{cases}$

$$\mu'_{1,1} = \frac{\mu_{1,1} \times 1/\|x'_1 - x_1\|^{2/(\partial-1)} + \mu_{1,2} \times 1/\|x'_1 - x_2\|^{2/(\partial-1)}}{1/\|x'_1 - x_1\|^{2/(\partial-1)} + 1/\|x'_1 - x_2\|^{2/(\partial-1)}}$$

$$= \frac{0 \times 1/1^{(\dots)} + 0 \times 1/1^{(\dots)}}{1/1^{(\dots)} + 1/1^{(\dots)}} = 0 \quad \rightarrow \mu'_{2,1} = 1$$

$$\mu'_{1,2} = \frac{\mu_{2,2} \times 1/\|x'_2 - x_2\|^{(2/(\partial-1))} + \mu_{2,3} \times 1/\|x'_2 - x_3\|^{2/(\partial-1)}}{1/\|x'_2 - x_2\|^{(\dots)} + 1/\|x'_2 - x_3\|^{(\dots)}}$$

$$= \frac{1 \times 1/3^{(2/(\partial-1))} + 0 \times 1/3^{(2/(\partial-1))}}{1/3^{(\dots)} + 1/3^{(\dots)}} \underset{\partial=2}{=} \frac{1/9}{1/9 + 1/9} = 0.5$$

$$U' = \begin{pmatrix} 0 & 0.5 \\ 1 & 0.5 \end{pmatrix}$$

→ x'_1 est dans la 2° classe

→ x'_2 est autant dans la 1ère classe que la 2° classe.

<u>Exercice R :</u>

* programmer Fuzzykmn (x_E, Data, k, ~~gamma~~ lambda) + Harden (Uj)

* programmer BestkFuzzy VFold

* Comparer les boxplots d'Erreur entre Fuzzy ("~~ne~~ fuzzyfié") et mon Fuzzy

* Tester les Valeurs de ~~gamma~~ lambda ($\in]1, +\infty[$)