La Classification Floue

1 Introduction

La classification floue est une généralisation de la classification classique (dite aussi classification exacte ou classification 'hard') dans le domaine des ensembles flous. L'avantage de l'utilisation de la notion du floue est la modélisation de l'incertitude et de l'ambigüité, tant qu'il y a encore du doute relativement à la classe à associer à certains individus. Ceci permet de retarder la classification tant qu'elle est incertaine. En effet, le processus de la classification floue crée une partition floue des données en affectant à chaque individu X_i un degré flou d'appartenance dans chacune des classes. Ainsi, l'appartenance n'est plus exclusive, vu que les individus ne sont plus affectés exclusivement à l'une ou l'autre des classes, mais on attribue à chaque individu X_i (avec i € {1, …, n) un degré d'appartenance continue $\mu_{ik} = \mu_{Ck}$ (X_i), dans l'intervalle [0, 1], et ceci à chacune des classes C_k (avec k € {1, …, C}, avec C est le nombre des classes), sachant que la somme des degrés d'appartenance d'un individu X_i à toutes les classes est égal à 1.

Notons que la classification floue est un cas particulier de la classification classique (dite aussi hard). En effet, dans le cas de la classification floue, $\mu_{ik} \in [0, 1]$ (incertitude), alors que dans le cas de la classification classique $\mu_{ik} \in \{0, 1\}$ (certitude).

Exemple : On veut détecter si un objet (un individu parmi ***n*** individus) appartient à l'une des trois catégories d'avions (hélicoptère (C1), avion de chasse (C2), avion civil (C3)) selon sa vitesse (***C=3***). Dans le cas classique, l'objet appartient exclusivement à (C1), (C2) ou (C3). Dans le cas de la logique floue, on définit pour chaque objet (individu) un degré d'appartenance à chacune des trois classes, tels que la somme des degrés d'appartenance d'un individu à toutes les classes est égal à 1.

Par exemple, μ_{C1}(objet) = 0.1, μ_{C2}(objet) = 0.7 et μ_{C3}(objet) = 0.2. Ceci revient à dire qu'on est certaine à 10% que l'objet en entrée appartient à la classe C_1, à 70% que l'objet en entrée appartient à la classe C_2, et à 20% que l'objet en entrée appartient à la classe C_1.

A la fin de la classification floue, on affecte chaque objet (individu) à la classe qui maximise le degré d'appartenance de l'objet, c'est-à-dire à la classe C_k tel que :

$$\mu_{C_k}(objet) = \max_j \mu_{C_j}(objet).$$

Dans la pratique, la méthode la plus fréquemment utilisée pour la réalisation de la classification floue, notamment en traitement des images, est l'algorithme FCM (Fuzzy C-Means). Cette méthode est une généralisation, dans le contexte de la logique floue, de l'algorithme de classification classique K-Means (dite Aussi C-Means, regroupement autour des centres mobiles).

L'algorithme FCM nécessite la reconnaissance du nombre de classes C au préalable et il cherche à générer des classes par un processus itératif en minimisant le critère de variabilité intra-classes J_m qui remplace l'inertie intra-classes.

Plus précisément, soient $X_1, \ldots, X_n$ les n individus et C le nombre de classes. Les valeurs des degrés d'appartenance sont regroupées dans une matrice U, avec :

$$U = (\mu_{ik})_{1 \leq i \leq n \ ; \ 1 \leq k \leq C} \text{ tels que } \mu_{ik} \in [0, 1] \text{ et } \sum_{k-1} \mu_{ik} = 1, \ \forall i = 1, \ldots, n.$$

Ainsi, l'algorithme FCM fait évoluer la partition (la matrice U) tout en minimisant J_m défini par :

$$J_m = \sum_{i=1}^{n} \sum_{k=1}^{C} (\mu_{ik})^m \|X_i - G_k\|_2^2$$

où $m > 1$ est un paramètre contrôlant le degré de floue (généralement $m = 2$). Le floue de la partition augmente avec ce coefficient car $(\mu_{ik})^m < \mu_{ik}$ puisque $\mu_{ik} \in [0, 1]$, mais au même temps, ce coefficient accentue les faibles niveaux d'appartenances et contribue donc à mieux séparer les classes.

Les centres des classes G_k sont définis comme suit :

$$G_k = \frac{\sum_{i=1}^{n} (\mu_{ik})^m X_i}{\sum_{i=1}^{n} (\mu_{ik})^m}, \forall k = 1, \ldots, C$$

Le centre de gravité G_k fait intervenir tous les individus pondérés par leurs degrés d'appartenances contrairement au cas classique où on fait intervenir seulement les individus de la même classe.

Ainsi, la solution au problème d'optimisation qui consiste à minimiser sachant J_m que

$$\sum_{k=1}^{C} \mu_{ik} = 1 \quad \forall i$$

est :

$$\mu_{ik} = \frac{\left(\frac{1}{d_{ik}}\right)^{\frac{2}{m-1}}}{\sum_{j=1}^{C}\left(\frac{1}{d_{ij}}\right)^{\frac{2}{m-1}}} \text{ avec } d_{ik} = \|X_i - G_k\|_2 \text{ et } d_{ik} \neq 0 \;;\; \mu_{ik} = 1 \text{ si } d_{ik} = 0.$$

L'algorithme FCM :

1. Input : le nombre de classes C, les centres G_k initiaux (initialisation aléatoire) ;
2. Initialiser de la matrice des degrès d'appartenances U ;
3. Calculer les centres des C classes $G_k^{(t)}$ en fonction de $U^{(t)}$ avec
$$G_k^{(t)} = \frac{\sum_{i=1}^{n}\left(\mu_{ik}^{(t)}\right)^m X_i}{\sum_{i=1}^{n}\left(\mu_{ik}^{(t)}\right)^m}$$
4. Calculer la matrice $U^{(t+1)}$ en fonction de $G_k^{(t)}$ avec
$$\mu_{ik}^{(t+1)} = \frac{1}{\sum_{j=1}^{C}\left(\frac{d_{ik}^{(t)}}{d_{ij}^{(t)}}\right)^{\frac{2}{m-1}}} \text{ avec } d_{ik}^{(t)} = \|X_i - G_k^{(t)}\|_2$$
5. On réitère les deux dernières étapes jusqu'au : Test d'arrêt $\mid J_m^{(t+1)} - J_m^{(t)} \mid < \varepsilon$.
6. Défuzzyfication : On affecte chaque individu à la classe pour laquelle il a le plus grand degrès d'appartenance.