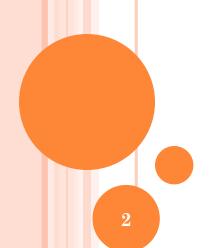


MACHINE LEARNING AND COMPUTATIONAL INTELLIGENCE MLCI

Dr. Fergani Baha.fergani@univ-constantine2.dz



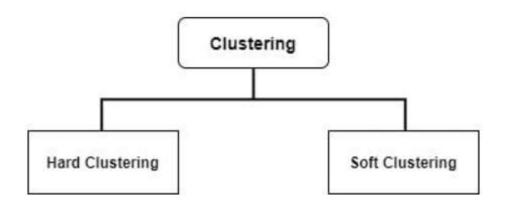
APPRENTISSAGE AUTOMATIQUE ET LOGIQUE FLOUE



LE CLUSTERING

- > Une technique d'apprentissage automatique non supervisée.
- Il divise la population en plusieurs clusters de sorte que les points de données d'un même groupe sont similaires les uns aux autres et que les points de données de différents groupes sont différents.
- Chaque cluster de la partition est défini par ses objets et son Centroïde.

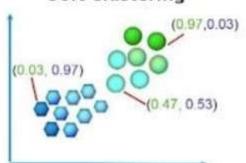
LE CLUSTERING



Hard Clustering



Soft Clustering



QU'EST-CE QUE LE CLUSTERING DUR?

- Dans le clustering dur, un élément appartient entièrement à un ensemble ou pas du tout.
- Un ensemble net est défini par une fonction de vérité bivalente qui n'accepte que les valeurs 0 et 1.
- **K-means** est un algorithme de clustering dur.
- > Il regroupe les points de données en k-clusters.

QU'EST-CE QUE LE SOFT CLUSTERING?

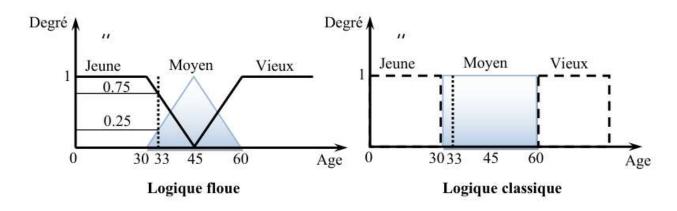
Clustering souple ≡ Soft clustering ≡ Clustering flou

Au lieu de placer chaque point de données dans des clusters distincts.

- > Une probabilité d'appartenance de ce point est attribuée aux clusters probables.
- Chaque point de données peut appartenir à plusieurs clusters avec un score de probabilité ou un degré d'appartenance.

Exemple:

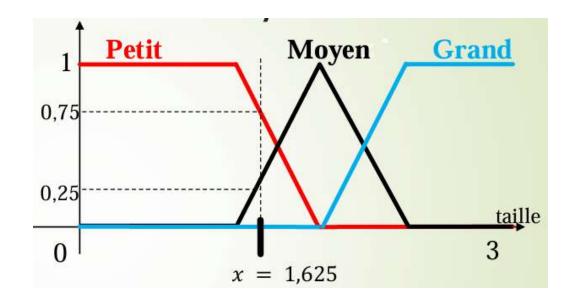
La classification des personnes selon leur âge en trois ensembles : jeune, moyen et vieux.



- Les limites entre ces trois groupes ne varient pas soudainement, mais progressivement.
- Une personne de 33 ans appartient à:
 - ✓ L'ensemble "jeune" avec une valeur μ =75.
 - ✓ L'ensemble "moyens" avec une valeur μ = 0.25.

Exemple:

- Mohammed mesure 1,625
- > «Mohammed est petit » avec un degré 75%.
- » «Mohammed est moyen » avec un degré 25%.
- > «Mohammed est grand » avec un degré 0%.

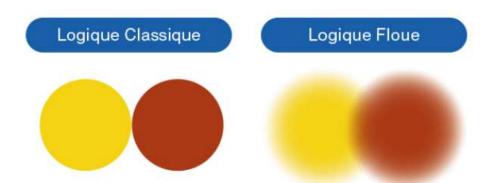


- Le terme d'ensemble flou apparaît pour la première fois en 1965 lorsque le professeur **Lotfi Zadeh**, de l'université de Berkeley aux USA, publie un article intitulé « Ensembles flous » (Fuzzy sets).
- La théorie des **ensembles flous** est une extension naturelle de la théorie classique des ensembles.
- Les ensembles flous peuvent être utilisés pour définir des expressions imprécises et ambiguës telles que "arbre haut" ou "voiture rapide".

- ➤ Un ensemble flou est déterminé par une fonction d'appartenance qui accepte toutes les valeurs intermédiaires entre 0 et 1.
- Les valeurs d'une fonction d'appartenance, appelées degrés ou grades d'appartenance, spécifient:
 - ✓ Précisément dans quelle mesure un élément appartient à un ensemble flou.

QU'EST-CE QUE LA LOGIQUE FLOUE?

- La *théorie de la logique floue* est basée sur la théorie des **ensembles flous**.
- > Une généralisation de la logique binaire.
- > Elle peut être utilisée pour la modélisation du raisonnement humain ou des langues naturelles.



AVANTAGES DE LA LOGIQUE FLOUE

- La théorie est simple et s'applique à des systèmes complexes.
- > Robustesse de la commande floue vis-à-vis des incertitudes.
- La maitrise du système à régler avec un comportement complexe.

INCONVENIENTS DE LA LOGIQUE FLOUE

- > Technique de réglage essentiellement empirique. Les performances dépendent de l'expertise.
- Le manque de directives précises pour la conception d'un réglage (choix des grandeurs à mesurer, détermination de la fuzzification, des inférences et de la défuzzification).
- L'approche artisanale et non systématique l'implémentation des connaissances de l'opérateur, est souvent difficile.

APPLICATIONS DE LA LOGIQUE FLOUE

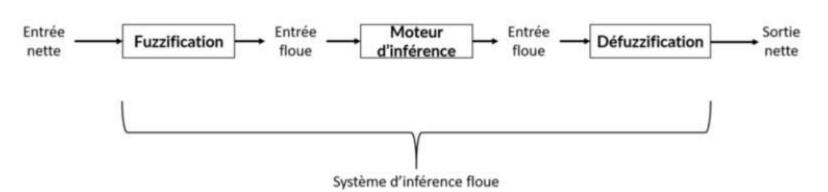
Domaine	Applications		
Médicale	-Cancer du sein		
	-Diabet		
	- Cardiotocographie		
	- Cancer de la thyroïde		
Traitement d'image	-Reconnaissance d'écriture		
	-Segmentation		
	-Reconnaissance des empreintes digitales, des visages		
	-La segmentation des images IRM		
	-La segmentation des images satellitaires		
	-Localisation d'objets dans des images		
Commerce en ligne	-La catégorisation des documents		
	-Le marketing		
	-la gestion de la clientèle		
	-Base de données clients en ligne		

Qu'est ce qu'un système intelligent flou: fis?

- > Un système qui intègre de l'expertise humaine.
- Il vise à automatiser le raisonnement d'experts humains face à des systèmes complexes.
- Il constitue une part importante de l'intelligence artificielle et du soft computing.
- > Un système intelligent qui se base sur la logique floue.

- > **But:** transformer les données d'entrée en données de sortie à partir de l'évaluation d'un ensemble des règles.
- > FIS utilise un ensemble de règles pour **prendre des décisions basées sur les données** d'entrée.
- Ces règles sont généralement exprimées sous la forme d'instructions "*if-then*", où la pièce "if" est les données d'entrée, et la pièce "alors" est la décision de sortie.
- Exemple: Si les données d'entrée sont "la température est élevée" et "l'humidité est faible", alors la décision de sortie peut être "allumer la climatisation".

- Un système à base de la logique floue est composé principalement de:
 - 1. La fuzzification (un fuzzificateur).
 - 2. Un moteur d'inférence.
 - 3. La défuzzification (un défuzzicateur).



La fuzzification (un fuzzificateur):

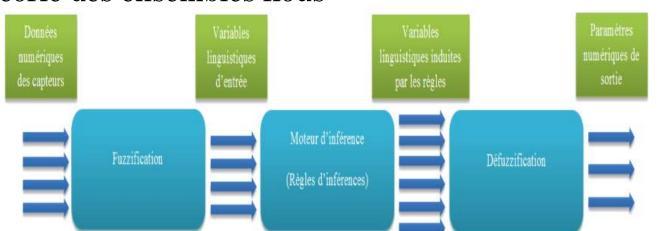
✓ Elle effectue la conversion de l'entrée numérique observée

$$u_{o} = (u_{o1}, u_{o2,..., u_{on}})^{T} \in U$$

✓ En un ensemble flou (défini dans U):

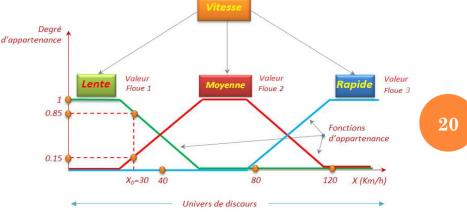
$$F_x = F_{xo} \times F_{x2} \times \dots F_{xn}$$

✓ Les données observées sont souvent numériques, mais dans le système flou la manipulation des données est basée sur la théorie des ensembles flous

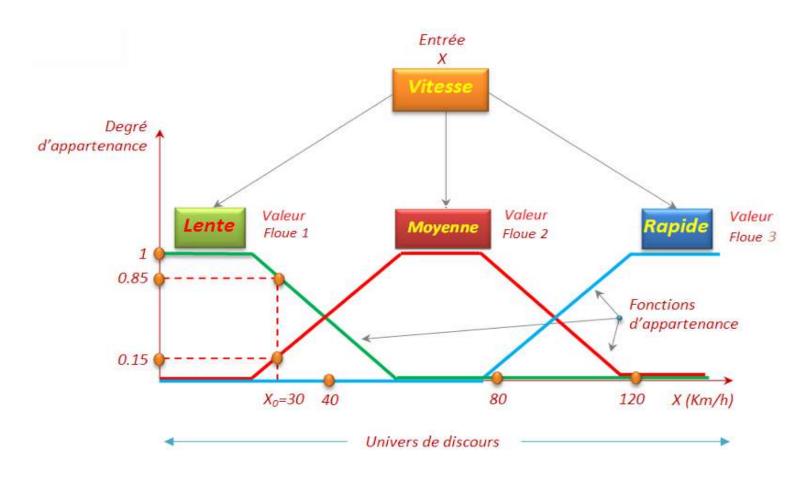


19

- La fuzzification (un fuzzificateur):
 - **But:** transformer une donnée numérique en variable linguistique.
 - Pour cela, le concepteur du système flou doit créer des fonctions d'appartenances.
 - Une fonction d'appartenance est une fonction qui permet de définir le degré d'appartenance d'une donnée numérique à une variable linguistique.



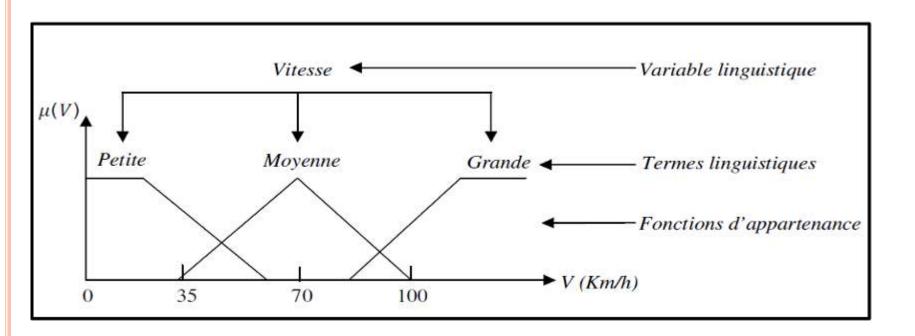
> La fuzzification (un fuzzificateur):



> La fuzzification (un fuzzificateur):

- La notion de variable linguistique permet de modéliser les connaissances imprécises ou vagues sur une variable dont la valeur précise est inconnue.
- Une variable linguistique, ou variable floue, est donc une variable dont les valeurs floues appartiennent à des ensembles flous pouvant représenter des mots du langage naturel.
- Le domaine sur lequel ces termes et ces variables sont définis, constitue **l'univers de discours**.

La fuzzification (un fuzzificateur):

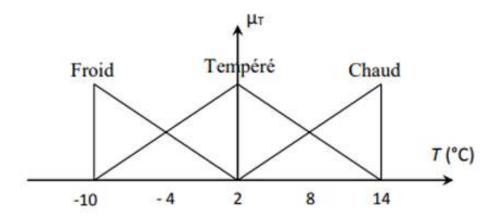


> La fuzzification (un fuzzificateur):

✓ Exemple:

Prenons la variable linguistique «Température»

- Univers de discours : Gamme de température de -10°C à 14°C.
- Variable linguistique : La température.
- Valeurs linguistiques : « Chaud » « Tempéré » « Froid ».

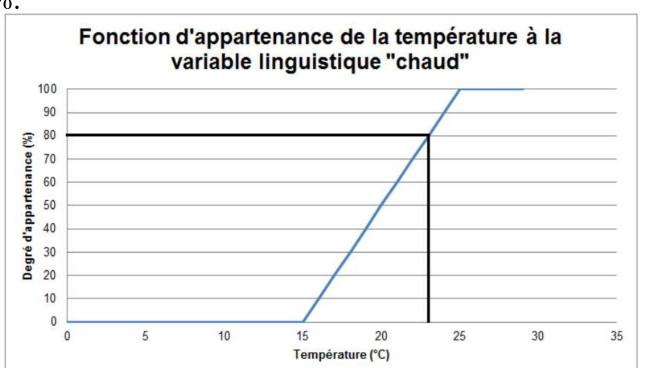


> La fuzzification (un fuzzificateur):

✓ Exemple:

- · Une température en degré Celcius provenant d'un capteur.
- On veut transformer cette donnée numérique en variable linguistique.
- On peut trouver plusieurs variables linguistiques qualifiant une température : chaud, froid, très froid, tempéré, très chaud, etc.
- Il suffit de créer une fonction d'appartenance de la température à chacune de ces variables. Créons la fonction d'appartenance de la température à la variable linguistique "chaud"

- La fuzzification (un fuzzificateur)
 - ✓ Si le capteur de température nous indique 23°.
 - ✓ La fuzzification nous donne: la température est chaude à 80%.

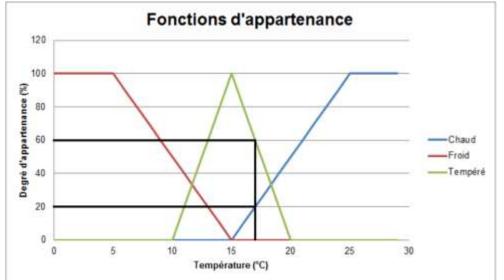


La fuzzification (un fuzzificateur)

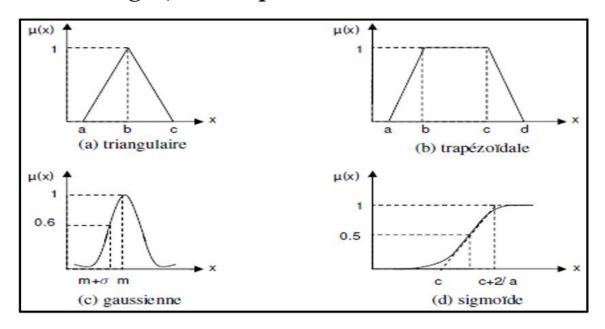
- On peut utiliser plusieurs variables linguistiques pour caractériser un seul type de données.
- Trois variables linguistiques peuvent être utilisées pour qualifier la température : chaud, froid et tempéré.

• Si le capteur nous renvoie 17°C, après fuzzification, la température sera chaude à 20%, tempérée à 60% et froide à

0%.

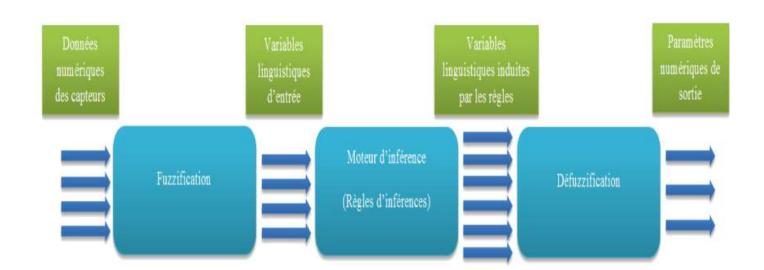


- La fuzzification (un fuzzificateur)
 - Les fonctions d'appartenances les plus courantes ont une forme de triangle, de trapèze ou de cloche.



Le moteur d'inférence

• Chaque règle du moteur d'inférence est écrite par le concepteur du système flou en fonction de connaissance qu'il possède.



Le moteur d'inférence

• Une règle doit être sous la forme:

Si

condition,

alors

conclusion.

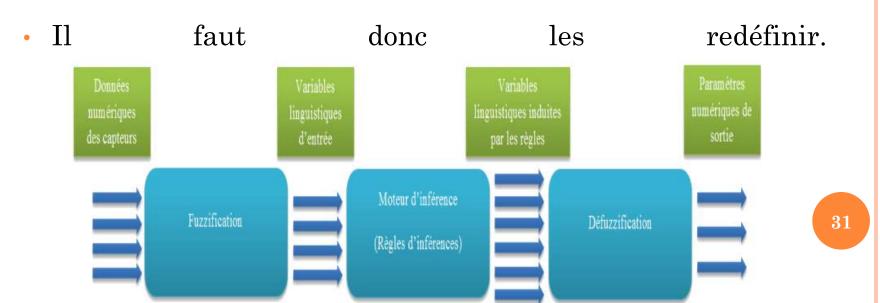
SI (température ≥ 25°) ALORS baisser le store

SI température EST bon ET éclairage EST moyen ALORS store EST à mi-hauteur

Si le feu est rouge	si ma vitesse est élevée	et si le feu est proche	alors je freine fort.
Si le feu est rouge	si ma vitesse est faible	et si le feu est loin	alors je maintiens ma vitesse.
Si le feu est orange	si ma vitesse est moyenne	et si le feu est loin	alors je freine doucement.
Si le feu est vert	si ma vitesse est faible	et si le feu est proche	alors j'accélère.

Le moteur d'inférence

- Par exemple, SI la vitesse est grande ET la distance au feu est courte ALORS freine fort, est une règle d'inférence valide.
- Les opérateurs de la logique classique (ET, OU) ne sont plus valables en logique floue.



L'opérateur ET:

- L'opérateur ET en logique floue correspond à l'intersection de deux ensembles flous.
- L'opérateur ET en logique floue peut être défini comme:
 - L'opérateur de minimalité : a ET b = min(a, b)
 - L'opérateur produit : a ET b = a.b
 - o Le ET flou: cet opérateur est un mixe entre l'opérateur de minimalité et la moyenne arithmétique.

$$a\ ET\ b = \gamma.min(a,b) + (1-\gamma).\frac{a+b}{2}$$

- La valeur de gamma est comprise entre 0 et 1 et qui doit être fixé par le concepteur du système flou.
- o Si elle vaut 1, on se retrouve avec l'opérateur de minimalité.

H A-8(x)

> Exemple:

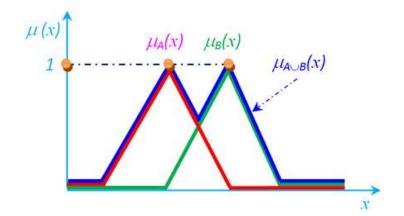
- Soit la règle: Si la vitesse est grande ET la distance au feu est courte ALORS freine fort.
- Si on sait que la vitesse est grande à 80% et que la distance au feu est courte à 20% alors
- En utilisant l'opérateur de minimalité → je vais freiner fort à 20%.
- En utilisant l'opérateur de produit → je freinerai fort à 16%.
- En utilisant l'opérateur ET flou avec un gamma à 0.5 :
 80% ET 20% = 0.5 min(80%, 20%) + 0.5 (80% + 20%) /2 = 35% → Je devrais donc freiner fort à 35%.

$$a ET b = \gamma .min(a,b) + (1-\gamma) .\frac{a+b}{2}$$

L'opérateur OU:

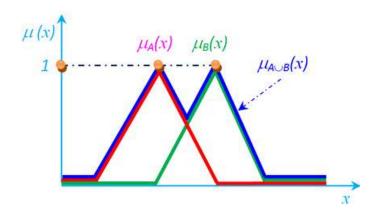
- Le OU en logique floue correspond à l'union de deux ensembles flous.
 - L'opérateur de maximalité : a OU b = max(a,b)
 - L'opérateur produit : a OU b = 1-(1-a).(1-b)
 - Le OU flou: une combinaison entre l'opérateur maximum et la moyenne arithmétique.

$$a OU b = \gamma . max(a,b) + (1-\gamma) . \frac{a+b}{2}$$



> Exemple:

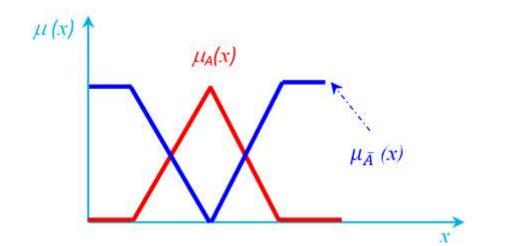
- Soit la règle: Si la vitesse est grande à 80% OU la distance au feu est courte ALORS freine fort à 20%.
 - o Si on utilise l'opérateur de maximalité → il faut freiner fort à 80%.



- ▶ L'opérateur NON en logique flou correspond à:
 - ✓ L'ensemble complémentaire et est défini:

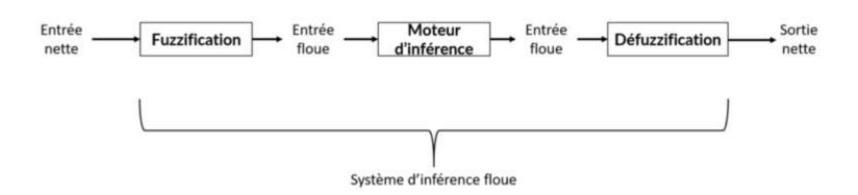
$$NON a = 1 - a$$

Le choix des opérateurs revient entièrement au concepteur du système flou.



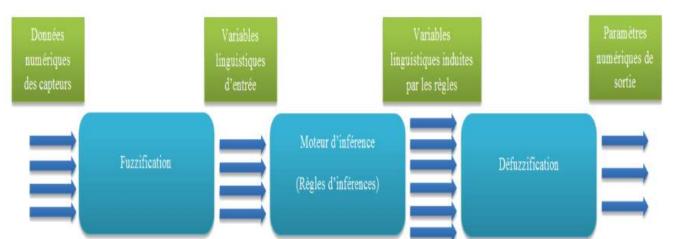
Defuzzification:

✓ La transformation floue en une information déterminée est la defuzzification (concrétisation).



> Defuzzification:

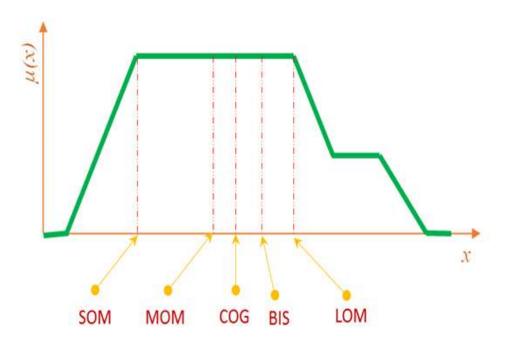
- Lors de la seconde étape, un tas de commandes a été généré sous la forme de variables linguistiques (une commande par règle).
- But: fusionner ces commandes et transformer les paramètres résultants en donnée numérique à l'aide d'un opérateur de la logique floue choisi par le concepteur du système.



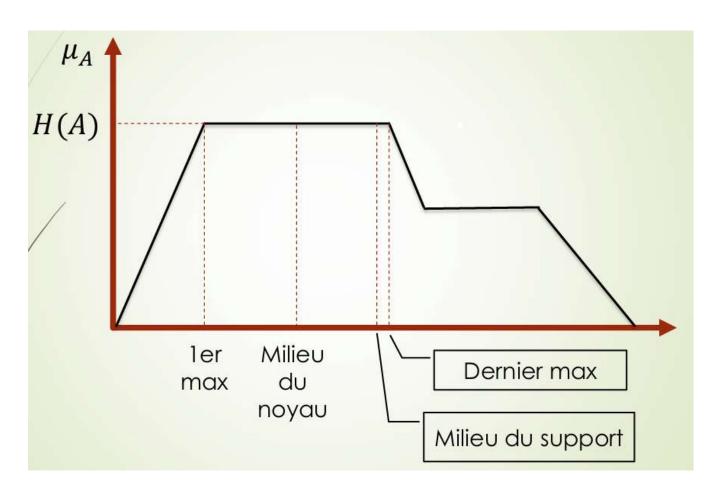
39

Defuzzification:

- ✓ COG: Centre de gravité.
- ✓ MOM: moyenne des maxima.
- ✓ LOM: plus grand maximum.
- ✓ SOM: plus petit maximum.



Defuzzification:



Defuzzification:

- Exemple:
 - o Soient trois règles qui génèrent la variable linguistique:
 - Accélère fortement à 20%.
 - Accélère fortement à 25%.
 - Accélère fortement à 35%.
 - o Si on utilise le OU logique en tant qu'opérateur de maximalité.
 - Il en résultera que la variable *accélère fortement* aura pour valeur finale: 35%.

En conclusion:

- o Pour réaliser un système flou, il faut:
 - Modéliser les données d'entrées en variables linguistiques grâce à des fonctions d'appartenances.
 - Dresser une liste de règles d'inférences qui représentent les connaissances que l'on a du système.
 - Il faut choisir quels sont les opérateurs logiques à utiliser et quel type de defuzzification on souhaite utiliser.
- Les performances d'un système flou seront à l'image des réglages que le concepteurs du système à bien voulu faire.
- Un système flou peut être très performant comme totalement inutile! (Par exemple, mon dernier exemple d'interpolation non linéaire n'est pas du tout performant car elle déforme les valeurs connaissances! Ainsi, l'interpolation de x=0.4 ne fait plus du tout 0.25 :s)

C-MEANS FLOU (FUZZY C-MEANS: FCM)

- > FCM a été développé par J.C. Dunn en 1973 et amélioré par J.C. Bezdek en 1981.
- Il permet à un élément de données d'appartenir à deux clusters ou plus.
- Le point se trouve dans chaque cluster jusqu'à un certain degré (en fonction de la fonction d'appartenance).
- > Par exemple, un point de données plus proche du centre aura une valeur d'appartenance plus élevée pour ce cluster par rapport à un centre de cluster plus éloigné de celui-ci.

- L'algorithme FCM tente de partitionner les éléments d'un ensemble fini de n Éléments $X=\{x_1,...,x_n\}$ en une collection de groupes flous C par rapport à un critère donné.
- > l'algorithme retourne une liste de \mathbf{c} Centres $\mathbf{C}=\{\mathbf{c}_1,...,\mathbf{c}_c\}$ et une matrice d'appartenance W.

$$W = w_{i,j} \in [0,1], \ i = 1, \ldots, n, \ j = 1, \ldots, c,$$

 \blacktriangleright Wij mesure le degrés d'appartenance de l'élément x_i au cluster dont le centre est c_j .

> FCM vise à minimiser une fonction objective

$$J(W,C) = \sum_{i=1}^n \sum_{j=1}^c w_{ij}^m \|\mathbf{x}_i - \mathbf{c}_j\|^2,$$

où:

$$w_{ij} = rac{1}{\sum_{k=1}^{c} \left(rac{\|\mathbf{x}_i - \mathbf{c}_j\|}{\|\mathbf{x}_i - \mathbf{c}_k\|}
ight)^{rac{2}{m-1}}}.$$

Initialisation:

- Initialisez le nombre de clusters C.
- Initialisez de manière aléatoire des centroïdes de cluster à partir de l'ensemble de données.
- Spécifiez le paramètre de flou (m) pour contrôler le degré de flou dans le clustering.

Mise à jour l'appartenance :

Calculez le degré d'appartenance de chaque point de données à chaque cluster.

Mise à jour du Centroid :

- Mettez à jour la valeur du centroïde.
- Recalculez les centroïdes du cluster en fonction des valeurs d'appartenance mises à jour.

□ Vérification de la convergence :

Répétez les étapes 2 et 3 jusqu'à ce qu'un nombre spécifié d'itérations soit atteint ou que les valeurs d'appartenance et les centroïdes convergent vers des valeurs stables.

INCONVENIENT DE FCM

Les degrés d'appartenance sont des degrés relatifs.

Si U_{ij} est le degré d'appartenance de X_i à la classe i Si U_{ij} est le degré d'appartenance de X_j à la classe i, la matrice $U_{C\times N}$ $[U_{ij}]$ est appelée matrice de C-partitions floues si et seulement si elle satisfait aux conditions:

L'appartenance d'un individu à une classe dépend de l'appartenance de cet individu aux autres clusters.

$$\begin{split} \forall i \in \{1..C\}, \forall j \ \in \{1 \dots N\} \begin{cases} & u_{ij} \in [0,1] \\ \\ & 0 < \sum_{j=1}^{N} u_{ij} < N \end{cases} \\ \forall j \in \{1 \dots N\} \sum_{i=1}^{c} u_{ij} = 1 \end{split}$$

Les points aberrants peuvent avoir des valeurs d'appartenance élevées et ils peuvent affecter de façon 50 significative l'estimation des centres des clusters.

K-MEANS / FCM

K-Moyenne	FCM
Initialisation du nombre de cluster k	1. Initialisation du nombre de cluster C
	Initialisation de ε , le degré de flou m (1,25 \leq m \leq 2)
Initialisation aléatoire des centres	2. Initialisation aléatoire des centres
MAJ des centres	3. Mettre à jour la matrice U des degrés d'appartenance :
$\overline{x_{kj}} = \frac{1}{ C_k } \sum_{i \in C_k} x_{ij}$	$U = \begin{bmatrix} u_{11} & \dots & u_{1k} & \dots & u_{1n} \\ u_{i1} & \dots & u_{ik} & \dots & u_{in} \\ u_{c1} & \dots & u_{ck} & \dots & u_{cn} \end{bmatrix} k = 1, \dots, n$ $i = 1, \dots, c$
	$u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}\right]^{-1}, \qquad u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{\ X_k - V_i\ }{\ X_k - V_j\ }\right)^{\frac{2}{m-1}}\right]^{-1}$
	4. MAJ le vecteur V des centres des clusters $v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m}$
Répéter jusqu'à convergence	5. Répéter les étapes 3 et 4 jusqu'à convergence $\ V^t - V^{t+1}\ < \varepsilon$ et le nombre d'itération

K-MEANS / FCM

FCM	K-Moyennes
Chaque point de données se voit attribuer un degré d'appartenance à chaque cluster, indiquant la probabilité ou la vraisemblance que le point appartienne à chaque cluster.	Chaque point de données est exclusivement attribué à un seul cluster, en fonction du centroïde le plus proche, généralement déterminé à l'aide de la distance euclidienne.
Il n'impose aucune contrainte sur la forme ou la variance des clusters. Il peut gérer des grappes de différentes formes et tailles, ce qui le rend plus flexible.	Il suppose que les grappes sont sphériques et ont la même variance. Ainsi, il peut ne pas bien fonctionner avec des grappes de formes non sphériques ou de tailles variables.
Il est moins sensible au bruit et aux valeurs aberrantes, car il permet des affectations de cluster souples et probabilistes.	Il est sensible au bruit et aux valeurs aberrantes dans les données

RÉSEAUX DE NEURONES FLOUS (NEURO-FUZZY)

RÉSEAUX DE NEURONES FLOUS (NEURO-FUZZY)

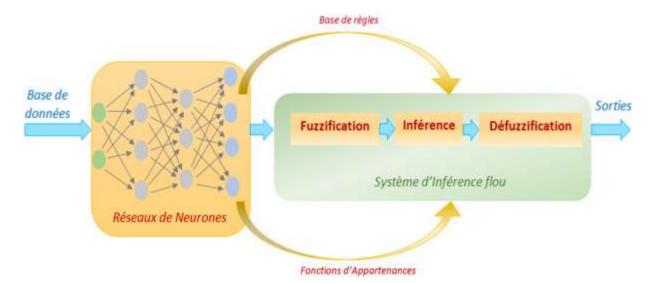
- □ Neuro-fuzzy correspond à la combinaison:
- > De réseaux de neurones artificiels.
- Des principes de la logique floue.
- L'hybridation neuro-fuzzy résulte en:
- Un système hybride intelligent.
- > Qui met en synergie ces deux techniques.
- En combinant le style de raisonnement des systèmes flous avec la structure d'apprentissage des réseaux de neurones.

Exemples:

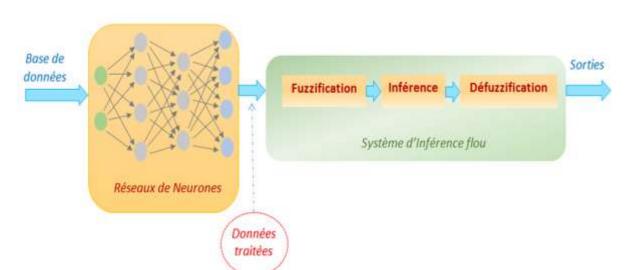
- En robotique, ces réseaux peuvent être utilisés pour contrôler le mouvement d'un robot dans un environnement incertain.
- Dans le traitement d'image, ces réseaux peuvent être utilisés pour reconnaître les objets dans une image qui peut être partiellement obscurcie ou difficile à identifier.
- Dans le traitement du langage naturel, ces réseaux peuvent être utilisés pour comprendre le sens des mots qui ont plusieurs interprétations

CLASSIFICATION DES RÉSEAUX DE NEURONES FLOUS

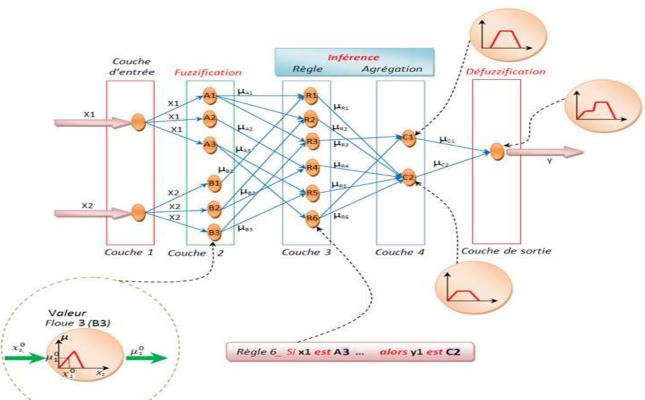
- > Système neuro-flou coopératif:
 - L'un des deux systèmes (flou ou neuronal) intervient une seule fois pour effectuer une tâche spécifique.
 - Les résultats obtenus du premier sont injectés et utilisés par le second système



- > Système neuro-flou concurrent:
 - L'un des deux systèmes (flou ou neuronal) intervient continuellement pour effectuer une tâche spécifique.
 - Les résultats obtenus du premier sont injectés et utilisés par le second système.

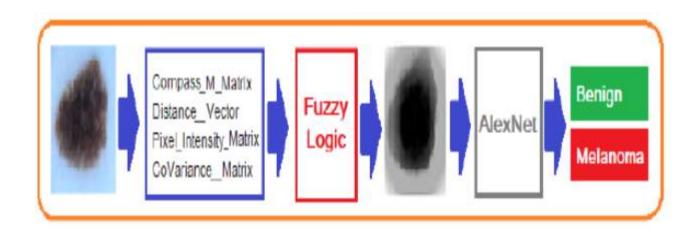


- > Système neuro-flou **hybride**:
 - Une combinaison homogène d'un réseau neuronal et un système d'inférence flou.



- > Système neuro-flou **hybride**:
 - ✓ Un réseau neuronal mené des paramètres flous (RN flou).
 - ✓ Un système flou dont les paramètres sont ajustés par apprentissage grâce au réseau de neurones (système flou neuronal).

- Un système neuro-flou profond combine deux paradigmes puissants:
 - ✓ Les réseaux neuronaux profonds.
 - ✓ La logique floue.



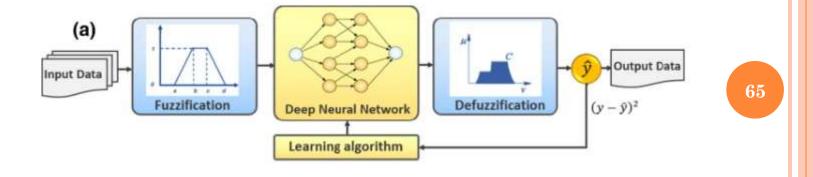
- Les systèmes neuro-flous profonds représentent une intégration de la capacité d'apprentissage profond pour apprendre de manière autonome des motifs et des représentations complexes avec la capacité d'interpretabilité et de raisonnement de la logique floue.
- > Cette fusion vise à exploiter les forces des deux approches pour résoudre des problèmes complexes du monde réel:
 - ✓ Ceux impliquant de l'incertitude.
 - ✓ De l'imprécision.
 - ✓ Des informations incomplètes.

Système neuro-flou profond

Exemples

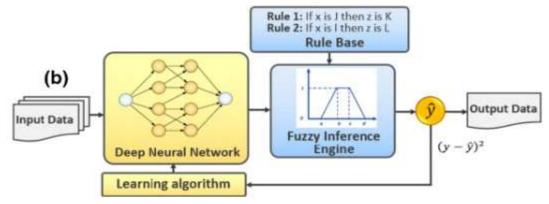
Conception structurelle coopérative:

- Le bloc d'interface floue convertit l'entrée nette en valeurs floues pour fournir un vecteur d'entrée à un DNN.
- > Le DNN est formé pour générer les sorties requises.
- La défuzzification des sorties est effectuée pour convertir la valeur floue en une valeur de sortie nette.

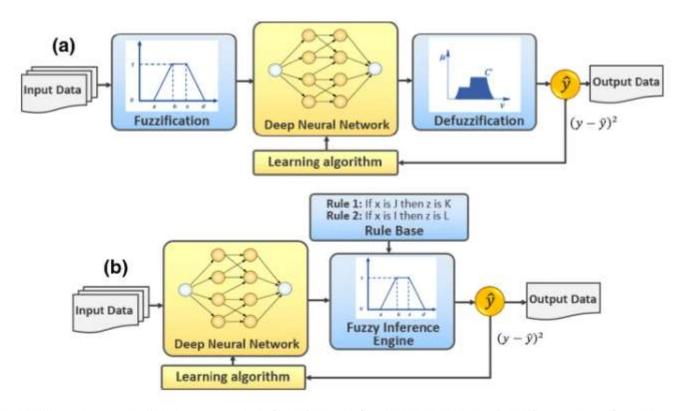


Conception structurelle coopérative:

- > Le mécanisme d'inférence floue est déterminé par un DNN.
- Les systèmes flous obtiennent les caractéristiques computationnelles de l'apprentissage offertes par un DNN.
- En retour, le DNN reçoit l'interprétation et la clarté de la représentation du système.



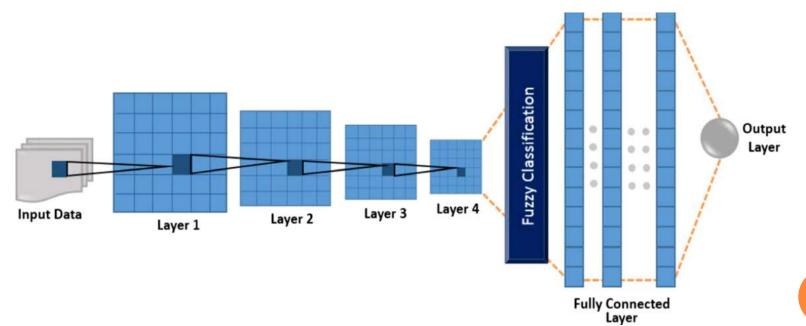
Conception structurelle coopérative:



Conception structurelle de DNFS coopératifs : un réseau neuronal profond flou et ${\bf un}$ réseau neuro-flou profond ${\bf b}$

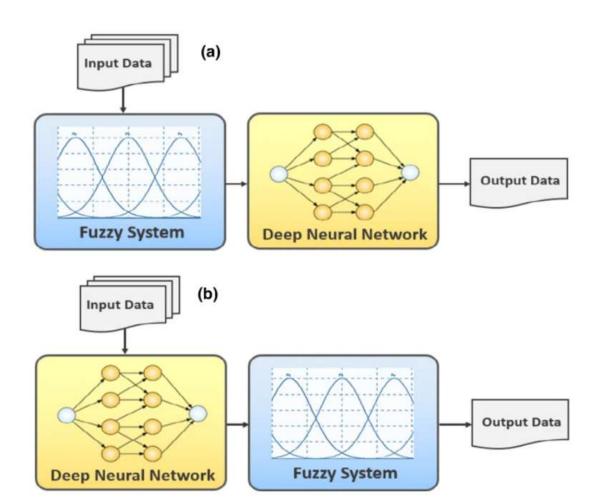
> Exemple:

Un exemple simple de DNFS coopératif est d'utiliser un CNN pour l'extraction de caractéristiques et en transférant les sorties de la couche convolutive finale



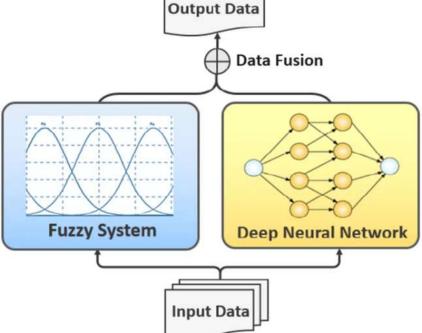
68

Conception structurelles séquentielle:



- Conception structurelle parallèle:
 - Les données sont traitées séparément des systèmes flous et d'un DNN.

Puis fusionnées pour obtenir le résultat final des données.



Merci pour votre attention