

Module IAS1122 : Intelligence artificielle

Méthodes de l'intelligence artificielle

Relwende Aristide YAMEOGO MD-MPH-PhD¹

¹Assistant hospitalo-universitaire en informatique médicale - UFR / Sciences de la Santé - Université Joseph KI-ZERBO

Master 2 IMSD
Année 2021 - 2022

7 - 10 Février 2022



Plan

- 1 Système expert
- 2 Logique floue
- 3 Système multi-agents
- 4 Machine Learning et Deep Learning
- 5 Données de l'IA

Définition

Système Expert

Les systèmes experts ou systèmes à base de règles (de production) sont des systèmes à base de connaissances dans lesquels les contenus de la base de connaissances se réfèrent au savoir et au savoir-faire des hommes de métier

- Mode de raisonnement basé sur la déduction
- Règle de production constituée d'une prémisse et d'une conclusion

Si condition(s) alors conclusion(s)

Modules du SE

- Trois modules séparés et interdépendants

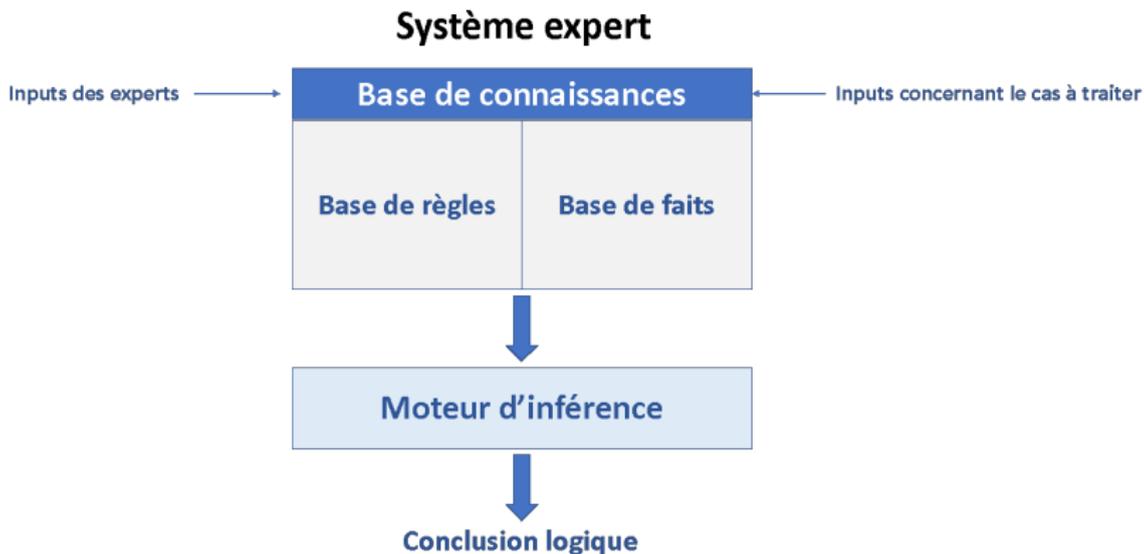


Figure 1 – Contenu d'un système expert

Chainages

Chainages

- le chaînage avant : la règle est déclenchée par la partie prémisses et la conclusion en résulte. On part des données disponibles dans la base de faits, et on teste pour chaque règle si elle peut s'appliquer ou non.
- le chaînage arrière : la règle est déclenchée par la partie conclusion ; pour valider la règle, il faut que la prémisses soit un fait connu.
- le chaînage mixte

Puissances

- Puissance du système expert
 - ▶ la quantité de connaissance traitée
 - ▶ la manière de les représenter : la base de faits et la base de règles
 - ▶ la manière de les exploiter : le moteur d'inférence
- Des moteurs de règles open source
 - ▶ CLIPS : moteur de règles dans le domaine public
 - ▶ Drools : distribué par RedHat
 - ▶ DTRules : moteur de règles en Java
 - ▶ Gandalf : moteur de règles tournant sur PHP

Définition

Logique floue

Le terme d'ensemble flou apparaît pour la première fois en 1965 lorsque le professeur Lotfi A. Zadeh, de l'université de Berkeley aux USA, publie un article intitulé « Ensembles flous » (Fuzzy sets) (ZADEH, 1965). C'est une technique de l'IA déterministe permettant de prendre des décisions. Elle permet ainsi d'avoir un comportement cohérent et reproductible en fonction des règles fournies.

Avantages

- Les avantages de la logique floue
 - ▶ formaliser et simuler l'expertise d'un opérateur ou d'un concepteur dans la conduite et le réglage d'un procédé
 - ▶ donner une réponse simple pour les procédés dont la modélisation est difficile
 - ▶ prendre en compte sans discontinuité des cas ou exceptions de natures différentes, et les intégrer au fur et à mesure dans l'expertise
 - ▶ prendre en compte plusieurs variables et effectuer de la « fusion pondérée » des grandeurs d'influence

Conditions d'utilisation

Pour que l'utilisation de règles floues puisse être envisagée, il faut impérativement qu'il existe une expertise, un savoir-faire humain. Les bases de règles floues n'apportent pas de solution quand personne ne sait comment fonctionne le système ou ne peut le piloter manuellement

Ensemble flou

■ Logique booléenne et logique floue

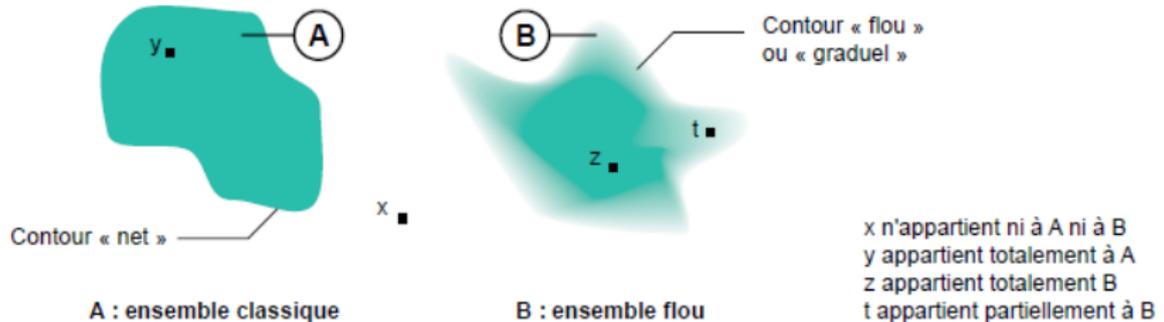


Figure 2 – Comparaison logique booléenne et logique floue

Fonction d'appartenance

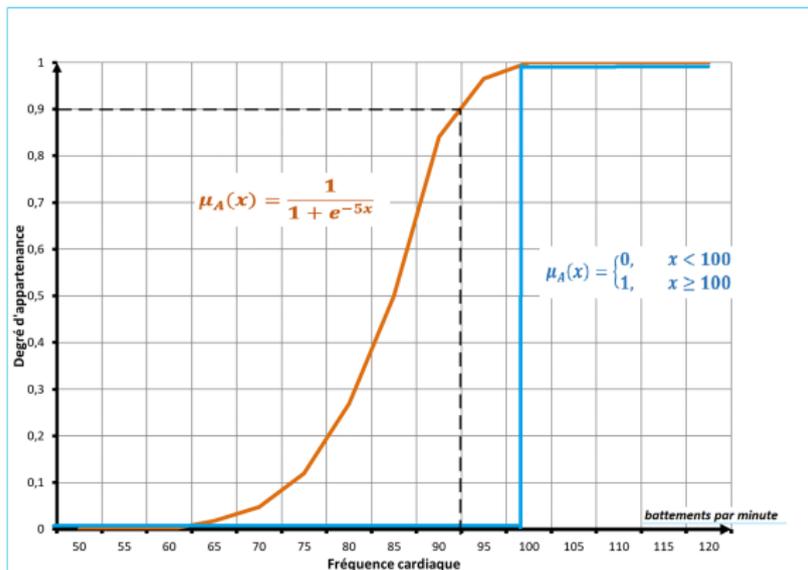


Figure 3 – Fonction d'appartenance d'un sous-ensemble flou et classique

Caractéristiques d'un ensemble flou

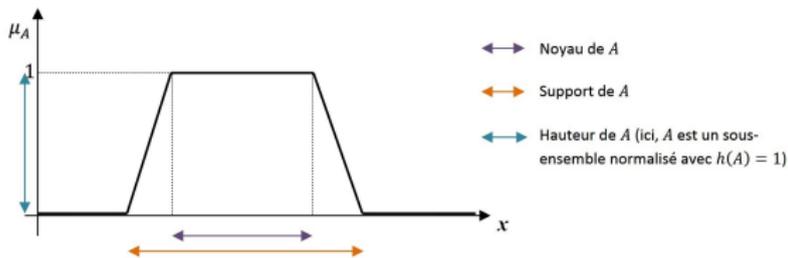


Figure 4 – Caractéristiques d'un ensemble flou

Types de fonction d'appartenance

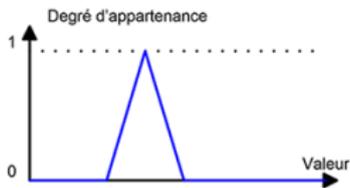


Figure 5 – Triangle

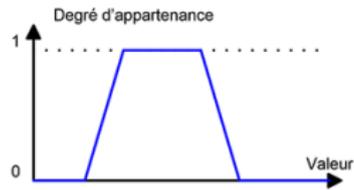


Figure 7 – Trapézoïde

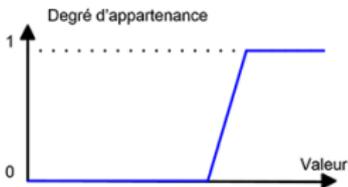


Figure 6 – Demi trapèze droite

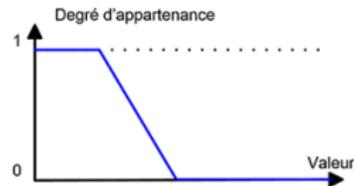


Figure 8 – Demi trapèze gauche

Types de fonction d'appartenance

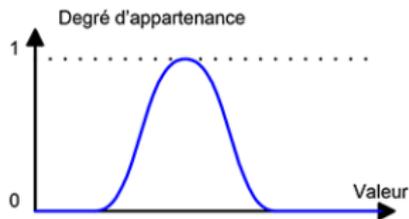


Figure 9 – Gaussienne

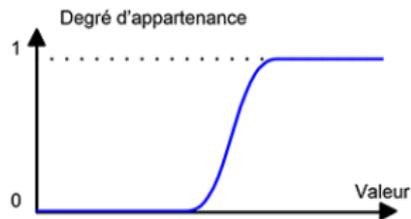


Figure 10 – Sigmoïde

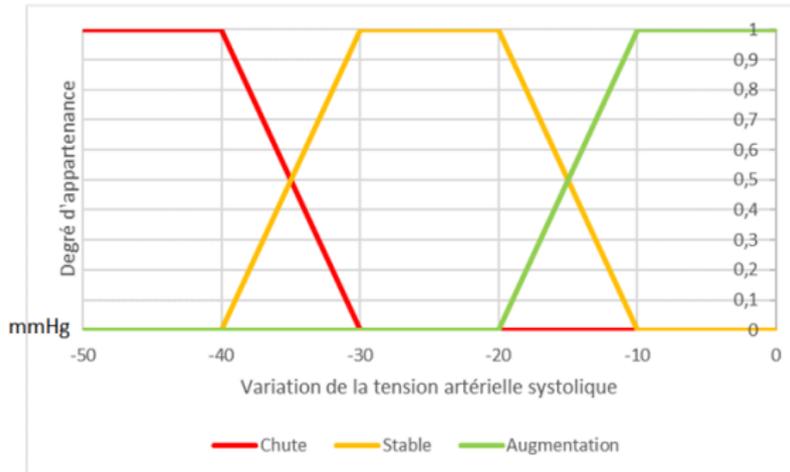
Définition

Définition

Une variable linguistique est représentée par un triplet (V, X, T_v) , dans lequel V est une variable (âge, fréquence cardiaque, tension artérielle) définie sur un ensemble de référence X (ensemble des nombres entiers, des réels . . .), sa valeur pouvant être m'importe quel élément de X .

« *Variable linguistique* » EST « *valeur linguistique* ».

Variables linguistiques



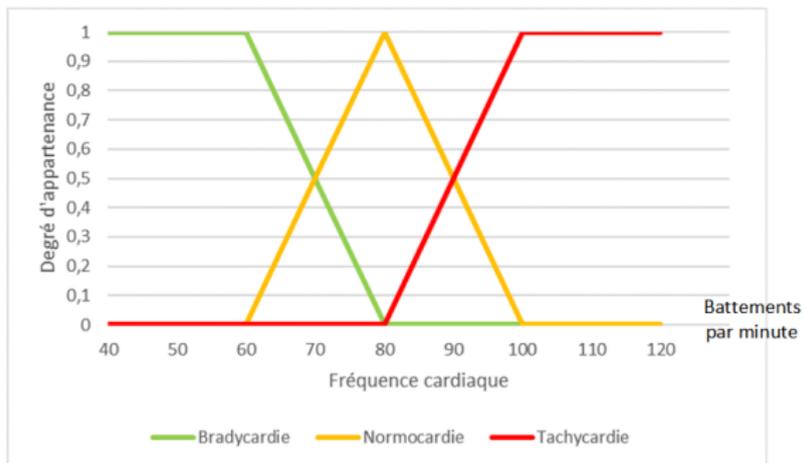
V = Variation de la tension artérielle systolique

$X = N$

$Tv = \{ chute, stable, augmentation \}$

Figure 11 – Tension artérielle

Variables linguistiques



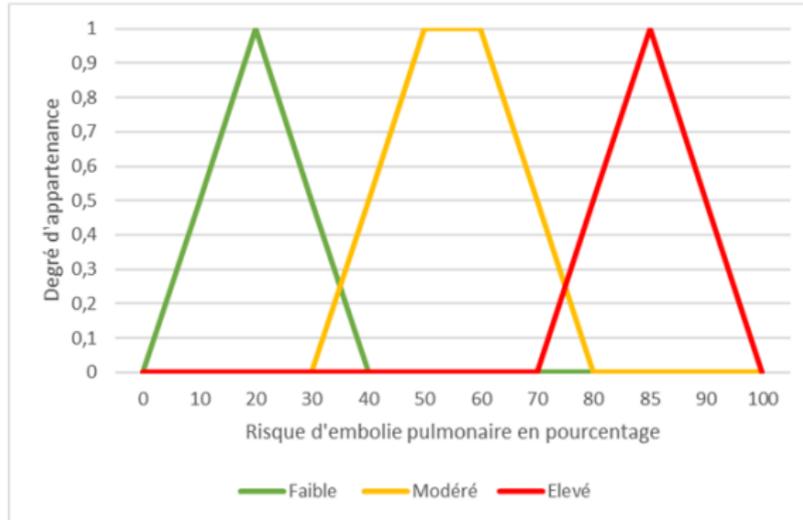
$V =$ Fréquence cardiaque

$X = N$

$T_v = \{ \text{bradycardie, normocardie, tachycardie} \}$

Figure 12 – Fréquence cardiaque

Variables linguistiques



V = Risque d'embolie pulmonaire
X = N
Tv = { faible, modéré, élevé }

Figure 13

Fuzzification

L'étape de fuzzification a pour but de transformer une donnée numérique en entrée en sous-ensemble flou. Pour cela, le concepteur du système flou doit utiliser des fonctions d'appartenances. Les différentes règles possèdent toutes une implication (la clause alors). Il va donc falloir exprimer à quel point la règle doit être appliquée en fonction des valeurs numériques mesurées.

SI Fréquence cardiaque **ET** Variation de la tension artérielle systolique **ALORS** Risque d'embolie pulmonaire

Fuzzification

Numéro de la règle	SI	Fréquence cardiaque	ET	Variation de la TAS	Alors	Risque d'embolie pulmonaire
1	Si	Tachycardie	Et	Chute	Alors	Élevé
2	Si	Tachycardie	Et	Stable	Alors	Modéré
3	Si	Tachycardie	Et	Augmentation	Alors	Modéré
4	Si	Normocardie	Et	Chute	Alors	Élevé
5	Si	Normocardie	Et	Stable	Alors	Modéré
6	Si	Normocardie	Et	Augmentation	Alors	Faible
7	Si	Bradycardie	Et	Chute	Alors	Élevé
8	Si	Bradycardie	Et	Stable	Alors	Faible
9	Si	Bradycardie	Et	Augmentation	Alors	Faible

Figure 14 – Règles d'inférence possibles

Fuzzification

Exemple

par exemple un patient qui a une fréquence cardiaque à 90 battements par minute et une variation de la tension artérielle systolique à -37 mmHg. Pour la variable linguistique « Fréquence cardiaque », la valeur appartient à normocardie dans 50% et tachycardie dans 50%. Pour la variable linguistique « Variation de la tension artérielle systolique », la valeur appartient à chute dans 70% et stable dans 30%. Ce qui nous permet de déduire quatre règles d'inférence (1, 2, 4 et 5).

Fuzzification

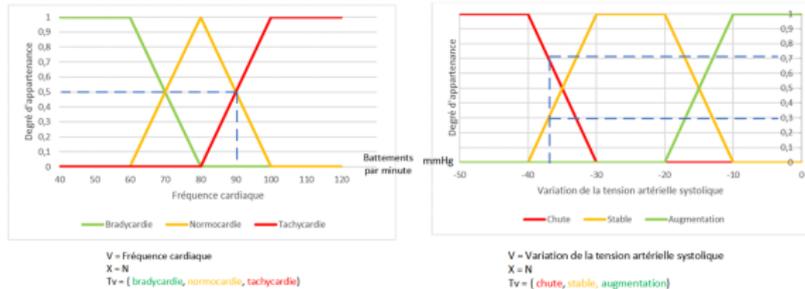


Figure 15 – Exemple d'inférence

Fuzzification

Opérateur d'implication de Mamdani

- l'agrégation qui est l'évaluation des conditions associées aux règles d'inférence, l'union étant caractérisée par l'opérateur de maximisation et l'intersection par la minimisation
- l'activation qui est l'évaluation de chaque inférence, et l'application de la relation d'inférence de Mamdani
- l'agrégation des conclusions qui est l'obtention de la fonction d'appartenance finale à partir du maximum des fonctions d'appartenance

Fuzzification

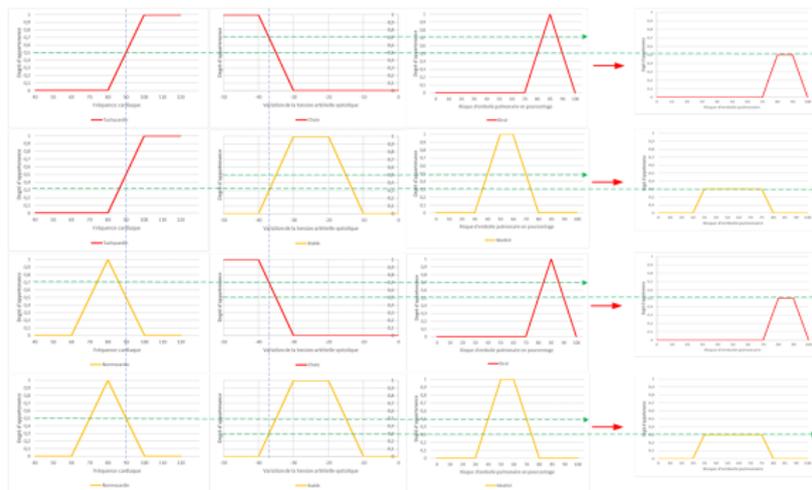


Figure 16 – Inférence de Mamdani

Défuzzification

Deux étapes :

- d'abord, il faut fusionner les variables linguistiques communes à l'aide d'un opérateur de la logique floue choisi par le concepteur du système. Si on a plusieurs règles d'inférences qui génèrent plusieurs valeurs de la même variable linguistique, on peut choisir un opérateur pour combiner les valeurs de la variable.

Défuzzification

Deux étapes :

- ensuite, on a des variables linguistiques qui caractérisent une seule et même donnée. Ces variables linguistiques possèdent chacune une fonction d'appartenance. Défuzzifier la donnée revient donc à trouver la meilleure valeur quantitative en fonction des fonctions d'appartenances des variables linguistiques.

Défuzzification

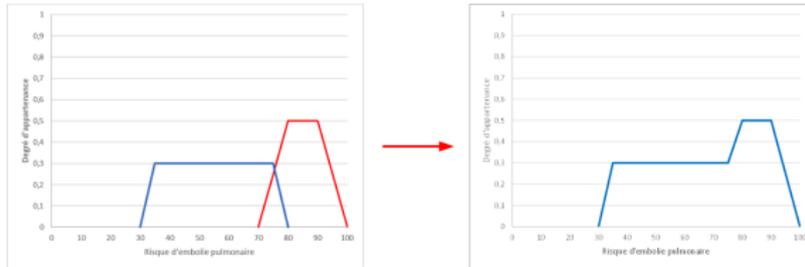


Figure 17 – Agrégation des conclusions

Défuzzification

Méthode de la moyenne des maximas

La méthode de la moyenne des maximas revient à prendre l'abscisse correspondant à la moyenne des abscisses ayant pour ordonnée la valeur maximale des fonctions d'appartenances.

Méthode du barycentre

La défuzzification par le barycentre, plus complexe, consiste à chercher le barycentre de la forme obtenue. Cette méthode consiste à calculer l'abscisse du centre de gravité de la surface formée par l'intersection de la courbe caractéristique de la fonction obtenue et l'axe des abscisses. Cette valeur retournée est la solution.

Défuzzification

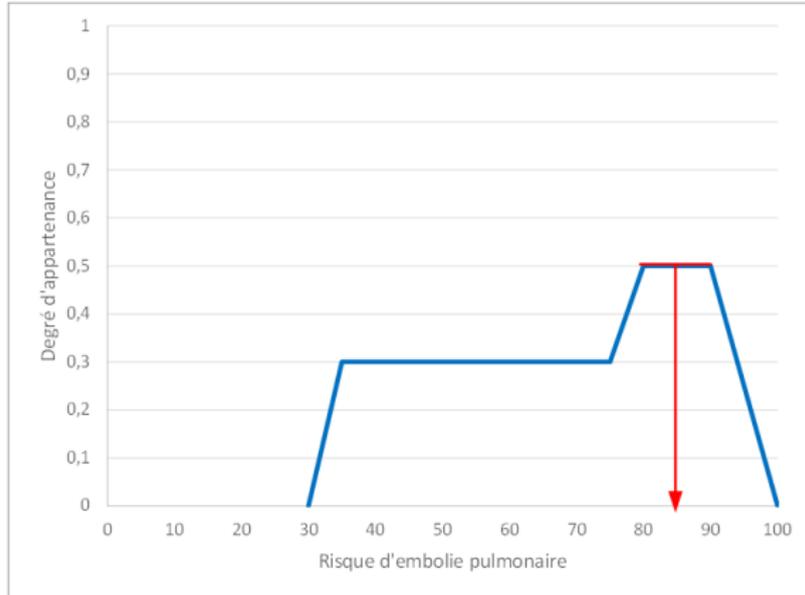


Figure 18 – Méthodes de la moyenne des maximas

Défuzzification

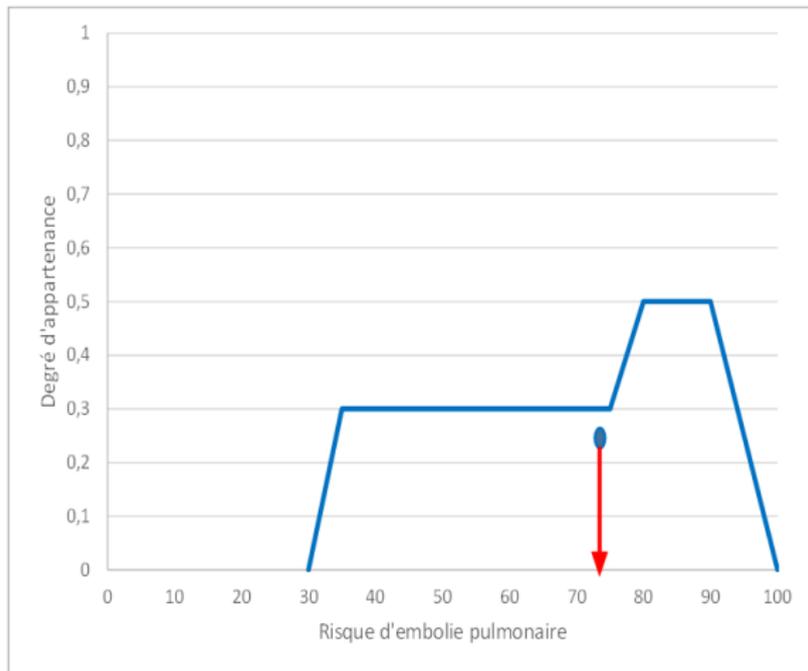


Figure 19 – Méthode du barycentre

Définition

Système multi-agents

Les concepts d'agents et de SMA sont apparus dans les années quatre-vingt. La motivation du domaine des SMA relativement récent est d'essayer de réaliser des entités artificielles (programmes, robots) qui simulent le comportement humain et la façon dont les hommes interagissent entre eux et avec leur environnement.

- Les disciplines en jeu dans les SMA
 - ▶ l'intelligence artificielle distribuée
 - ▶ les systèmes distribués
 - ▶ le génie logiciel

Intelligence artificielle distribuée

Intelligence artificielle distribuée

L'intelligence artificielle distribuée (IAD) se définit comme la branche de l'IA qui s'intéresse à la modélisation de comportement intelligent d'un ensemble d'agents opérant collectivement et de façon décentralisée pour parvenir à un objectif global. Elle propose une distribution de l'expertise sur un groupe d'agents, non soumis à un contrôle centralisé, devant être capable de travailler et d'agir dans un environnement commun et de résoudre des problèmes

Système distribué

Système distribué

Un système distribué est un ensemble d'entités autonomes de calcul (ordinateurs, PDA, processeurs, processus, processus léger etc.) interconnectées et qui peuvent communiquer. La technologie distribuée permet au SMA de se focaliser sur les couches supérieures du logiciel et de mettre en évidence leur plus-value-réelle : la prise de décision distribuée et coopérative, l'autonomie, l'adaptabilité,

Mise en oeuvre

- Pour une mise en oeuvre de l'intelligence sociale
 - ▶ un environnement, correspondant aux problèmes que l'on cherche à résoudre
 - ▶ des objets fixes ou non, étant des obstacles ou des points d'intérêt, dont la présence est facultative
 - ▶ des agents, aux comportements simples
- Le but de l'algorithme n'est jamais réellement codé, la solution va émerger de l'interaction de tous ces éléments entre eux

Définition

- Agent vient de « agri » : un agent est donc une entité susceptible d'action
- Un agent sera dit rationnel si le choix de ses actions est asservi à la satisfaction de ses besoins, ce qui sous-entend implicitement, qu'un agent rationnel possède des désirs, des souhaits, des buts
- Agent rationnel = un système intentionnel

Définition

Système intentionnel

La notion de système intentionnel sert à décrire des entités naturelles ou artificielles trop complexes pour que l'on ne maîtrise toutes les déterminations, et à qui l'on prête des « intentions », au sens précédent, c'est-à-dire des affects, des préférences, des buts, ... relatifs aux objets du monde, afin d'en prévoir le comportement

Définition

Définition d'un agent selon Ferber

un agent est « une entité autonome, réelle ou abstraite, qui est capable d'agir sur elle-même et sur son environnement, qui, dans un univers multiagent, peut communiquer avec d'autres agents, et dont le comportement est la conséquence de ses observations, de ses connaissances et des interactions avec les autres agents »

Définition d'un agent selon Jennings, Sycara et Wooldridge

un agent est un système informatique, situé dans un environnement, et qui agit d'une façon autonome et flexible pour atteindre les objectifs pour lesquels il a été conçu

Caractéristiques d'un agent

- Principales caractéristiques d'un agent
 - ▶ l'autonomie : un agent contrôle son état interne et ses actions. Il est capable d'agir sans l'intervention d'un tiers
 - ▶ la réactivité : un agent perçoit son environnement et réagit à ses changements
 - ▶ la pro activité ; un agent tend par lui-même à atteindre des buts (explicites ou non, plus ou moins élaborés)
 - ▶ la sociabilité : un agent interagit avec d'autres agents du système quand la situation l'exige afin de compléter ses tâches ou aider ces agents à compléter les leurs
 - ▶ la capacité de répondre à temps : l'agent doit être capable d'élaborer une réponse dans les temps requis
- les caractéristiques des agents permettent de définir une typologie des agents en fonction de leur capacité de raisonnement

Différents types d'agents

Agent réactif

les agents réactifs, souvent qualifiés de peu intelligents, possèdent une représentation très simplifiée de leur environnement. Leurs capacités consistent à réagir uniquement en mode stimulus/action vu comme une forme de communication (ou perception).

Différents types d'agents

Agent cognitif

les agents cognitifs possèdent une représentation partielle mais sophistiquée de leur environnement, ont des buts explicites et sont capables de planifier leur comportement, de mémoriser leurs actions passées, de communiquer par envoi de messages ou via des langages d'interactions élaborés, de négocier

Les agents cognitifs peuvent être

- intentionnels (dotés d'attitudes intentionnelles telles que les croyances, les désirs et les intentions),
- rationnels (agissant selon une rationalité donnée telle que la rationalité économique), normés (agents évoluant dans un système doté de normes sociales)

Différents types d'agent

- **Éléments de mesure de la rationalité**
 - ▶ la mesure de performance qui définit le degré de succès d'un agent
 - ▶ l'historique de la perception d'un agent
 - ▶ la connaissance de l'agent à propos de son environnement
 - ▶ les actions qu'un agent peut effectuer
- Un agent rationnel idéal est défini comme : « pour tout historique possible de sa perception, cet agent effectuera l'action dont les connaissances dont il dispose permettent de penser qu'elle maximise sa performance »

Espace d'un SMA

- Espace d'un SMA : trois axes obligatoires
 - ▶ **l'axe cognitif** fait référence aux aptitudes cognitives d'un agent telles que l'apprentissage, la prise de décision, la planification autonome qui permettent aux agents de s'adapter aux changements potentiels de leur environnement
 - ▶ **l'axe d'interaction** fait référence aux mécanismes de communication entre agents qui leur permettent d'échanger de l'information, de partager des objectifs et de coopérer dans le cadre de la résolution des conflits et des négociations
 - ▶ **l'axe de concurrence** signifie que les agents peuvent s'exécuter en parallèle et mener des activités concurrentes. Ceci permet aux agents de distribuer la résolution des tâches et de l'optimiser, ou enfin de concourir à l'affectation/résolution de tâches.

Algorithmes couramment utilisés

Algorithme de meutes

L'algorithme de meutes : à partir de quelques règles simples, il est possible de simuler des comportements de meutes ou de groupes. Craig Reynolds crée ainsi en 1986 les boids, des créatures artificielles évoluant en groupe

Algorithmes couramment utilisés

Optimisation par colonie de fourmie

L'optimisation par colonie de fourmis est directement inspirée du fonctionnement des fourmis éclaireuses. Le but est donc de trouver une solution optimale grâce à la stigmergie . Cette technique a été créée par Marco Dorigo en 1992.

Stigmergie

En biologie, la stigmergie est un mécanisme de coordination indirecte entre les agents. Le principe est que la trace laissée dans l'environnement par l'action initiale stimule une action suivante, par le même agent ou un agent différent.

Algorithmes couramment utilisés

Système immunitaire artificiel

Les systèmes immunitaires artificiels s'inspirent des systèmes naturels des vertébrés (comme le nôtre). En effet, plusieurs cellules collaborent pour déterminer le soi et le non soi, et attaquer ce qui a été déterminé comme étranger (virus, bactéries, champignons, voire poisons). Les systèmes immunitaires artificiels font donc évoluer, dans un environnement donné, différents agents de défense. Chacun connaît un ensemble de menaces, qu'il sait détecter (et combattre si nécessaire). Ces agents peuvent être créés aléatoirement au départ

Algorithmes couramment utilisés

Automate cellulaire

L'automate cellulaire : son environnement est une grille régulière. Chaque agent est situé sur une des cases de la grille et ne peut en bouger. Il possède différents états possibles (par exemple des couleurs). À chaque itération, il va pouvoir changer d'état en suivant quelques règles simples toujours basées sur les états de ses voisins proches. La grille va donc évoluer au cours du temps

Applications pratiques

■ Utilisations

- ▶ Contrôle du trafic aérien
- ▶ Surveillance continue des patients en USIC
- ▶ SADM

■ Difficultés

- ▶ le codage
- ▶ le choix de l'environnement
- ▶ le choix des agents et surtout de leurs caractéristiques

Définitions

Machine Learning

Le machine learning est une méthode d'analyse de données permettant d'automatiser le développement de modèle analytique. Par le biais d'algorithmes capables d'apprendre de manière itérative, le machine learning permet aux ordinateurs de découvrir des insights cachés sans être programmés pour savoir où les chercher..

Définitions

Deep Learning

Le deep learning, ou apprentissage profond, fait référence à un type d'intelligence artificielle, qui lors d'analyses complexes, les caractéristiques essentielles du traitement des données ne seront plus identifiées par un traitement humain dans un algorithme préalable, mais directement par l'algorithme de deep learning.

Définitions

Système d'information autonome

Le système d'information autonome est un modèle de système de décision dont les opérations consistent à coordonner de nombreuses décisions d'actions que le système doit prendre à chaque instant t , en tenant compte des informations endogènes (sa propre activité et régulation) et exogènes : les informations provenant de l'environnement (entrées) et ses actions décidées pour tenter d'agir sur l'environnement (sorties)

Définitions

Algorithme

Le Petit Robert 2014 définit le mot « algorithme » comme « suite finie, séquentielle de règles que l'on applique à un nombre fini de données, permettant de résoudre des classes de problèmes semblables » ; « ensemble des règles opératoires propres à un calcul ou à un traitement informatique », « Calcul, enchaînement des actions nécessaires à l'accomplissement d'une tâche »

Prédiction

La prédiction consiste à faire l'hypothèse d'un événement futur en se fiant aux données observables du passé ou du présent

Mode d'apprentissage

- Apprentissage supervisé

Apprentissage supervisé

L'apprentissage supervisé qui est un type de machine learning dont le but est d'apprendre à faire des prédictions, à partir d'une liste d'exemples étiquetés, c'est-à-dire accompagnés de la valeur à prédire. Les étiquettes servent de modèles d'enseignement et supervisent l'apprentissage. le système exploite les exemples et acquiert la capacité à les généraliser ensuite sur de nouvelles données de production.

Mode d'apprentissage

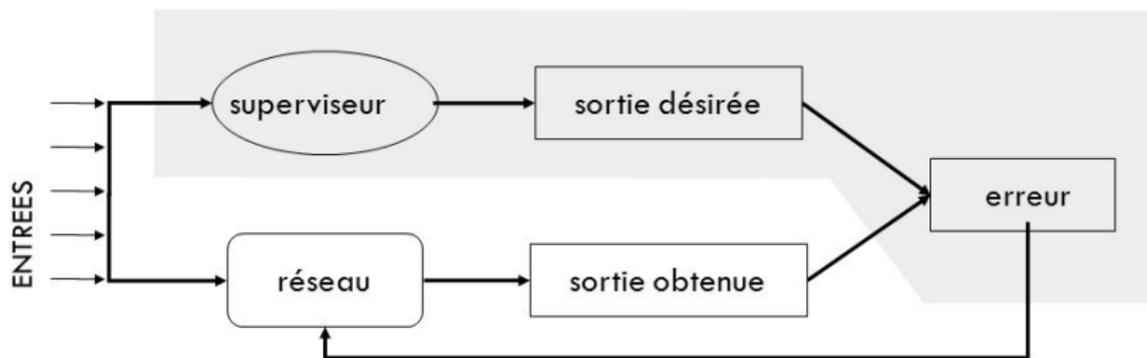


Figure 20 – Apprentissage supervisé

Mode d'apprentissage

- Apprentissage non supervisé

Apprentissage non supervisé

L'apprentissage non supervisé où les données ne sont pas étiquetées. Il faut alors modéliser les observations pour mieux les comprendre. Le système ne dispose pas de retour explicite sur ses actions

Mode d'apprentissage

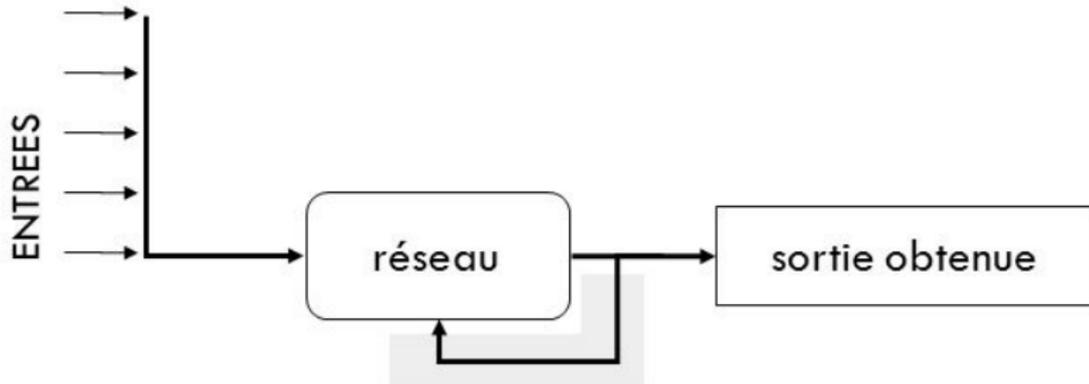


Figure 21 – Apprentissage non supervisé

Mode d'apprentissage

Apprentissage par renforcement

L'apprentissage par renforcement est un mode d'apprentissage où le système peut interagir avec son environnement et accomplir des actions. Le système apprend par des mécanismes de récompenses et de punitions. L'apprentissage consiste dans ce cas à définir une stratégie permettant d'obtenir le meilleur résultat possible. C'est une forme d'apprentissage supervisé incrémental qui utilise des données arrivant au fil de l'eau pour modifier le comportement du système

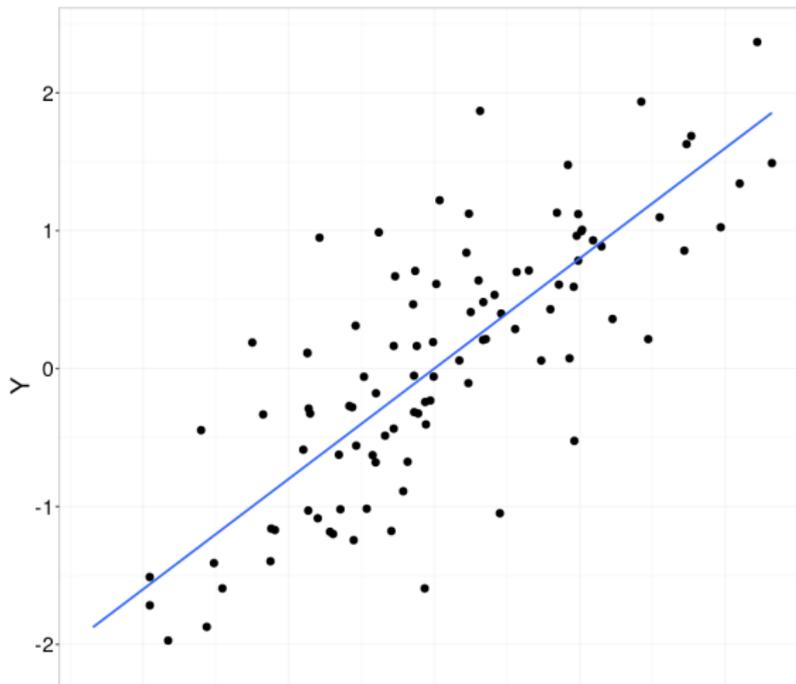
Différents types de méthodes

Régression linéaire

- Une régression permet de prédire une valeur numérique y en fonction d'une valeur x à partir d'un jeu d'entraînement constitué de paires de données (x, y) . x peut être une seule valeur numérique ou une combinaison de valeurs (vecteur ou matrice de données)
- Cet algorithme cherche à établir, sous forme d'une droite, une relation entre une variable expliquée et une variable explicative. Par exemple, prédire la taille d'une personne (variable expliquée) en fonction de son âge (variable explicative).

Différents types de méthodes

Régression linéaire



Différents types de méthodes

Régression linéaire

- La normalisation ou la mise à l'échelle (scaling) est une étape importante et va donc consister à faire en sorte que la moyenne de chaque série d'observations soit égale à 0, que la variance et l'écart-type soient égaux à 1.
- Cette méthode est également appelée centrage de réduction. Cette méthode permet de faire la prédiction grâce à la méthode de descente de gradient ou à la méthode des moindres carrés).

Différents types de méthodes

Régression linéaire

Méthode de descente de gradient

La méthode de descente de gradient : Lorsqu'un système est en phase d'apprentissage, il commet des erreurs. Le taux d'erreur diminue au fur et à mesure de l'apprentissage, mais il se peut qu'à un moment donné l'erreur augmente pour à nouveau diminuer et atteindre un niveau d'erreur plus bas que le précédent qui est le niveau optimal d'apprentissage.

Différents types de méthodes

Régression linéaire

Méthode des moindres carrés

La méthode des moindres carrés consiste à trouver la droite qui minimise l'écart entre les points théoriques et les points pratiques (mesures réelles). Le point théorique qui correspond au point pratique est celui situé sur la droite de régression à la même abscisse. La méthode consiste à prendre la somme des écarts au carré, et à les minimiser

Différents types de méthodes

Régression polynomiale

Il est parfois difficile de trouver une droite pouvant passer parmi les points de la série d'observations de façon optimale. Cependant, il est parfois possible de trouver un lien entre les variables à l'aide d'une courbe. C'est ce que permet la régression polynomiale en ajoutant des plis à la courbe à l'aide d'éléments appelés polynômes

Différents types de méthodes

Régression polynomiale

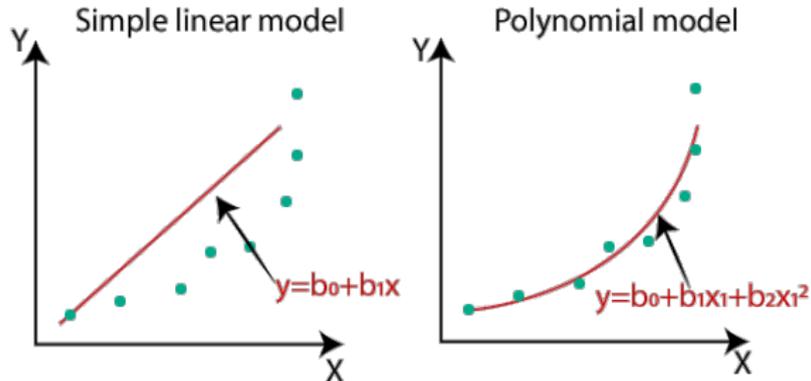


Figure 23 – Régression polynomiale

Différents types de méthodes

Régression logistique

Elle utilise, quant à elle, une fonction logistique encore appelée sigmoïde ou courbe en S. Ce type d'algorithme est à appliquer dans des problèmes de classification. La régression logistique a pour but d'isoler les effets de chaque variable, c'est-à-dire d'identifier les effets résiduels d'une variable explicative sur une variable d'intérêt, une fois pris en compte les autres variables explicatives introduites dans le modèle

Différents types de méthodes

Régression logistique

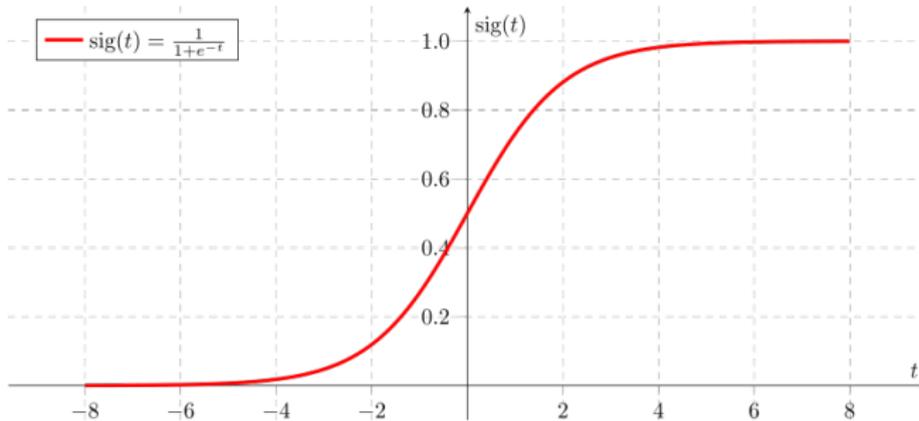


Figure 24 – Régression logistique

Différents types de méthodes

Régression logistique

Deux types de compromis

- le biais du modèle est la capacité du modèle à représenter la réalité
- la variance de l'apprentissage est l'erreur induite par la sensibilité du modèle aux fluctuations de l'échantillon d'apprentissage).

Différents types de méthodes

Arbre de décision

Les arbres de décision exploitent des critères discriminants, comme dans un moteur de règles. Ils permettent de classifier en effectuant une série de tests. Chaque nœud interne correspond à un test sur la valeur de l'un des attributs d'entrée, et les branches partant de ce nœud sont étiquetées avec les valeurs possibles de cet attribut

Différents types de méthodes

Arbre de décision

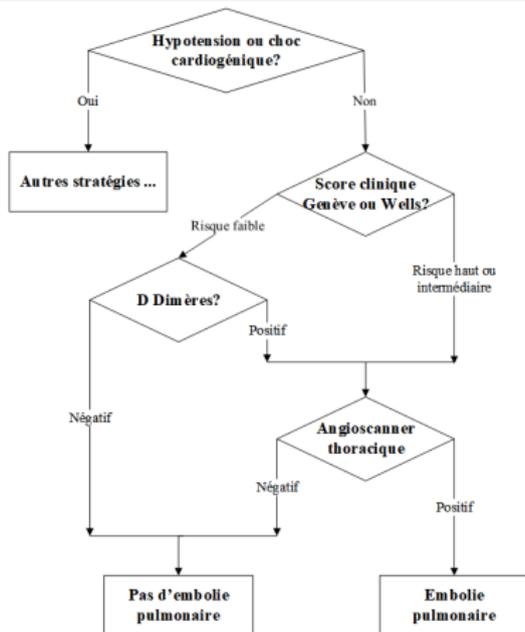


Figure 25 – Arbre de décision

Différents types de méthodes

Random forest ou forêt aléatoire

- Pour chaque test, on partage l'ensemble des données en sous-ensembles et on effectue la moyenne des valeurs de ces sous-ensembles en guise de prédiction
- Son fonctionnement consiste à apprendre en parallèle sur plusieurs arbres de décisions construits aléatoirement et entraînés sur des sous-ensembles contenant des données différentes
- Chaque arbre propose alors une prédiction et la prédiction finale consiste à réaliser la moyenne de toutes les prédictions
- La complexité dans l'utilisation de cet algorithme est de trouver le bon nombre d'arbres à utiliser

Différents types de méthodes

Machine à vecteurs de support (Support Vector Machine (SVM))

- Le SVM est un algorithme puissant utilisé aussi bien dans les cas de classification que de régression
- Son objectif est de déterminer une frontière afin de séparer les observations en groupes distincts tout en maximisant la marge de séparation
- Les observations les plus proches de la frontière sont appelées des vecteurs de supports
- Les SVM mémorisent tous les exemples d'apprentissages et peuvent avoir besoin de tous les stocker

Différents types de méthodes

Machine à vecteurs de support (Support Vector Machine (SVM))

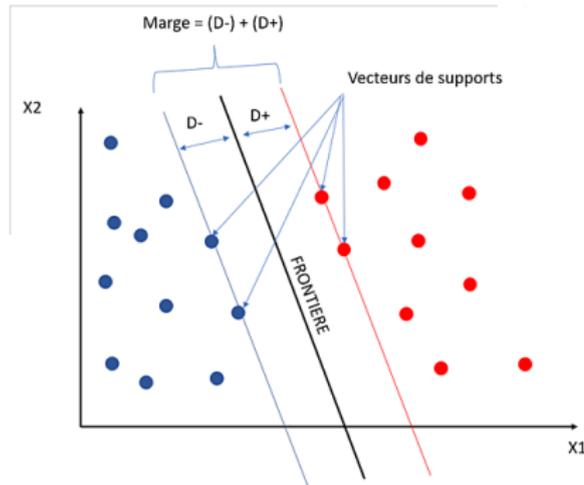


Figure 26 – Machine à vecteurs de support

Différents types de méthodes

K-Nearest Neighbours (k-NN)

- L'algorithme K-Nearest Neighbours est un algorithme de classification visant à déterminer pour une observation donnée son groupe d'appartenance à partir du groupe d'appartenance de ses K plus proches voisins. K étant le nombre de voisins à considérer
- La fonction de décision s'exprime en fonction des données observées et non pas comme une formule analytique fonction des variables. La procédure d'apprentissage consiste à stocker les données du jeu d'entraînement et ne comporte aucun calcul ; mais peut devenir un facteur limitant dans le cadre des grandes bases de données

Différents types de méthodes

Naive Bayes

- Naive Bayes est un algorithme de classification fondé sur les probabilités.
- Naive Bayes se base sur le Théorème de Bayes fondé sur les probabilités conditionnelles, c'est-à-dire la détermination de la probabilité qu'un évènement se produise en fonction d'un évènement qui s'est déjà produit
- À chaque hypothèse de départ, on associe une probabilité. L'observation d'une ou de plusieurs instances peut modifier cette probabilité. On peut parler de l'hypothèse la plus probable au vu des instances observées.
- Les probabilités bayésiennes présupposent l'indépendance des attributs utilisés

Différents types de méthodes

K-Moyennes (K-Means)

- L'algorithme des K-Moyennes (K-Means) est un des algorithmes de clustering
- Le clustering, le partitionnement ou la segmentation automatique est une méthode d'apprentissage non supervisé qui permet à partir d'un jeu de données non labellisé d'identifier des groupes de données proches les unes des autres, les clusters de données
- Les méthodes de clustering permettent d'identifier les paramètres discriminants de ces différents segments

Différents types de méthodes

K-Moyennes (K-Means)

- Après avoir indiqué à l'algorithme le nombre de clusters à trouver, celui-ci tente par itérations successives de déterminer des centroïdes (un par cluster) autour desquels il est possible de regrouper les données
- Ces regroupements s'effectuant en calculant la distance de chaque observation par rapport à un point central de regroupement appelé centroïde et permettant ainsi de classer les observations en plusieurs groupes de façon automatique
- La méthode des K-Means cherche à minimiser la variance intracluster afin de créer des clusters homogènes.

Différents types de méthodes

Réseaux de neurones

Perceptron

Le perceptron, réalisé en 1957 par Frank Rosenbalt, est un algorithme d'apprentissage supervisé. Le perceptron sous sa forme la plus simple est également nommé neurone formel et a pour but de séparer des observations en deux classes (ou groupes) distinctes à condition que ces données soient linéairement séparables. Son rôle est donc de classifier

Différents types de méthodes

Réseaux de neurones

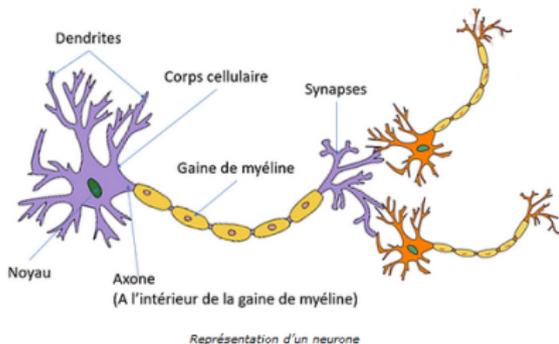


Figure 27 – Neurone

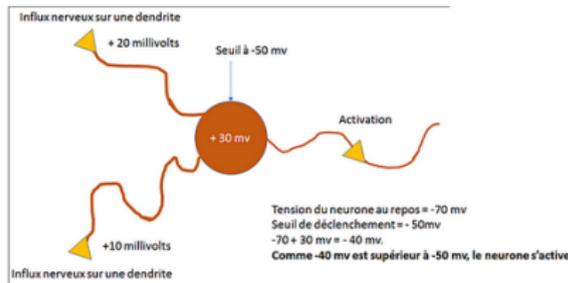


Figure 28 – Activation neuronale

Différents types de méthodes

Réseaux de neurones

Un neurone reçoit des entrées et fournit une sortie, grâce à différentes caractéristiques :

- des poids accordés à chacune des entrées, permettant de modifier l'importance de certaines entrées par rapport aux autres
- une fonction d'agrégation, qui permet de calculer une unique valeur à partir des entrées et des poids correspondants
- un seuil (ou biais), permettant d'indiquer quand le neurone doit agir
- une fonction d'activation, qui associe à chaque valeur agrégée une unique valeur de sortie dépendant du seuil

Différents types de méthodes

Réseaux de neurones

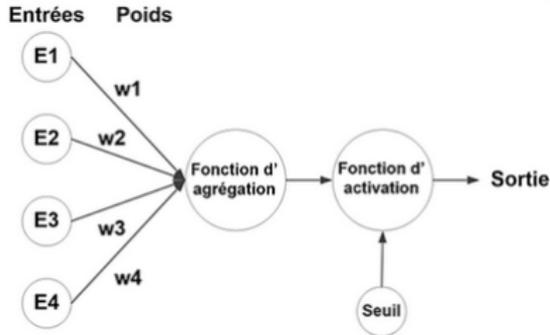


Figure 29 – Principe de fonctionnement

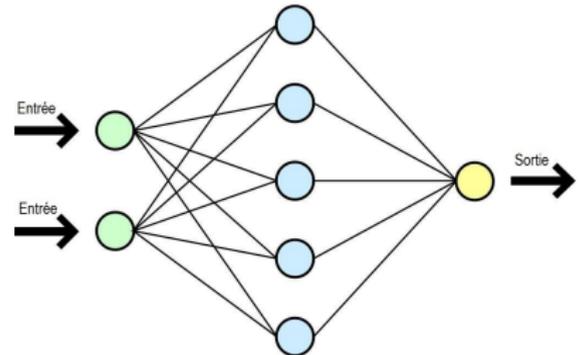


Figure 30 – Réseau de neurone à une couche

Différents types de méthodes

Réseaux de neurones

■ Fonction d'agrégation

Le but de cette fonction est de transformer l'ensemble des entrées et des poids en une seule valeur, qui sera utilisée pour la fonction d'activation. Les deux plus courantes sont :

- ▶ la somme pondérée : on va simplement faire la somme de toutes les entrées (E) multipliées par leur poids (w)
- ▶ Le calcul des distances : on va comparer les entrées aux poids (qui sont les entrées attendues par le neurone), et calculer la distance entre les deux. La distance est la racine carrée de la somme des différences au carré

Différents types de méthodes

Réseaux de neurones

■ Fonction d'activation

L'activation du neurone artificiel est déclenchée en fonction d'un seuil comme le neurone biologique

- ▶ La première étape dans la détermination de l'activation ou non d'un neurone artificiel est de réaliser la somme des poids synaptiques avec les valeurs d'entrées
- ▶ Une fois cette somme réalisée, on utilise une fonction d'activation qui va permettre de déterminer si le neurone doit s'activer ou non. Cette activation sera en fait la prédiction du neurone.

Différents types de méthodes

Réseaux de neurones

■ Fonction d'activation

Pour calculer cette prédiction, il existe plusieurs fonctions utilisables :

- ▶ la fonction de seuil binaire : cette fonction retourne une valeur égale à 0 quand le seuil d'activation défini n'est pas atteint ou 1 quand le seuil est atteint.
- ▶ la fonction sigmoïde dont les changements de valeurs entre 0 et 1 sont plus progressifs. La formule mathématique de cette fonction donne naissance à la courbe en S.
- ▶ la fonction tangente hyperbolique est une fonction sigmoïde. Elle présente cependant l'avantage que pour toute valeur négative celle-ci reste ou devient fortement négative

Différents types de méthodes

Réseaux de neurones

■ Fonction d'activation

Pour calculer cette prédiction, il existe plusieurs fonctions utilisables :

- ▶ la fonction ReLU (Rectified Linear Unit, unité de rectification linéaire) : La fonction ReLU pallie ce problème de saturation des fonctions sigmoïdes et hyperboliques via un fonctionnement assez simple. Si la valeur issue de la somme pondérée est inférieure à 0, celle-ci prend la valeur 0, sinon elle prend la valeur de la somme calculée
- ▶ la fonction softMax attribue des probabilités décimales à chaque classe d'un problème à plusieurs classes. La somme de ces probabilités décimales doit être égale à 1. Ce type de fonction d'activation est généralement utilisé dans un réseau de neurones multicouche et dans le cas de classifications multi classes

Différents types de méthodes

Réseaux feed-forward

Les réseaux de type feed-forward ou à couches permettent de dépasser les limites des perceptrons

- Dans les réseaux feed-forward l'information ne peut aller que des entrées aux sorties, sans revenir en arrière
- Le nombre de couches cachées et le nombre de neurones de chaque couche sont à choisir par l'utilisateur (généralement en faisant des tests sur plusieurs valeurs)
- On parle de deep learning quand le nombre de couches est important (à partir de trois ou quatre couches cachées).
- Dans le cas des réseaux feed-forward, il existe un apprentissage possible par rétro propagation du gradient

Différents types de méthodes

Réseaux feed-forward

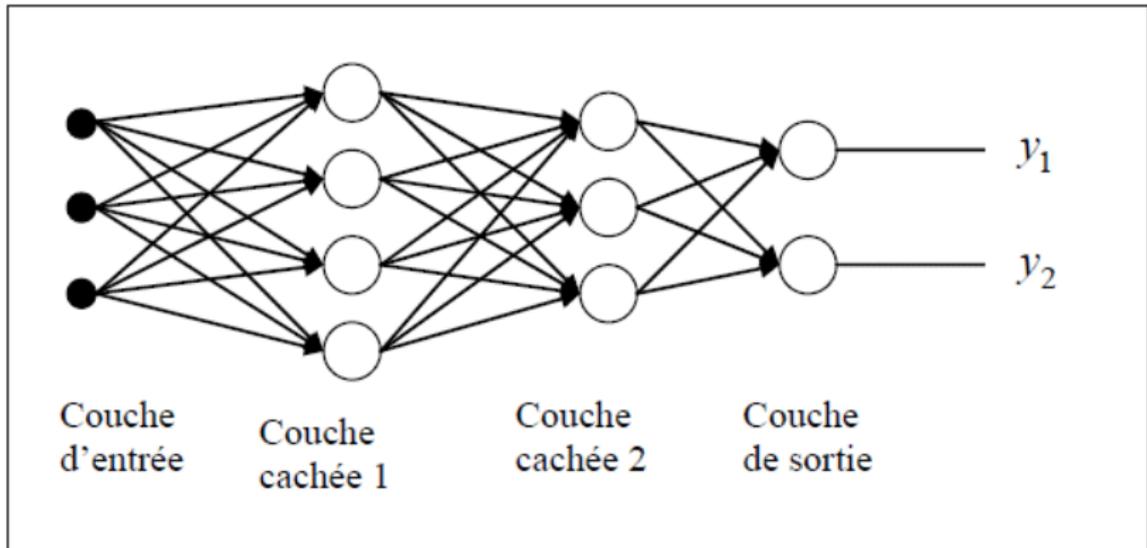


Figure 31 – Réseau de neurones multicouches

Différents types de méthodes

Réseaux feed-forward

Retro propagation

Le principe de l'apprentissage par rétro propagation est de corriger tout d'abord les poids entre les neurones de sortie et les neurones cachés, puis propager l'erreur en arrière et corriger les poids entre les neurones cachés et les entrées.

Différents types de méthodes

réseaux de neurones convolutifs ou Convolution Network Neurone (CNN)

Ils sont très adaptés pour travailler sur des images et des vidéos, le traitement de langage naturel et les systèmes de recommandations. Ces réseaux sont composés de multiples couches avec des rôles différents

- les couches de convolution, qui traitent des parties de l'image (extracteurs de caractéristiques)
- les couches de pooling (qui combinent les sorties des couches de convolution pour en détecter des fonctionnalités de plus haut niveau)
- les couches plus classiques de classification (souvent avec la fonction ReLU comme fonction d'activation)

Différents types de méthodes

réseaux de neurones convolutifs ou Convolution Network Neurone (CNN)

- Un traitement correctif non linéaire est appliqué entre chaque couche pour améliorer la pertinence du résultat. L'ensemble des sorties d'une couche permet de reconstituer une image intermédiaire dite de « carte de caractéristique » (feature map), qui sert de base à la couche suivante
- Ces réseaux sont cependant très durs à entraîner en partant de 0. Généralement, on utilise un réseau connu pré entraîné, et on ne fera que de l'entraînement sur notre jeu de données sur quelques générations. Cette utilisation d'un réseau existant s'appelle transfert learning ou apprentissage par transfert.

Différents types de méthodes

réseaux de neurones convolutifs ou Convolution Network Neurone (CNN)

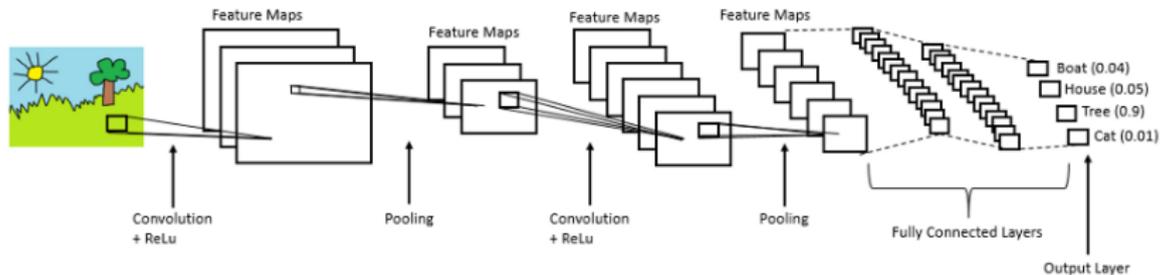


Figure 32 – Architecture d'un reseau de neurones convolutifs

Différents types de méthodes

Algorithmes génétiques

Les algorithmes génétiques appartiennent à la famille des algorithmes évolutionnistes. Ils sont utilisés pour effectuer des optimisations sur des problèmes complexes afin d'obtenir une solution proche de l'optimal. Les algorithmes génétiques utilisent la notion de sélection naturelle et l'appliquent à une population de solutions potentielles au problème donné. Les algorithmes génétiques ont déjà une histoire relativement ancienne puisque les premiers travaux de John Holland sur les systèmes adaptatifs remontent à 1962

Différents types de méthodes

Algorithmes génétiques

En entrée, un algorithme génétique dispose :

- d'une population initiale à n solutions ; chaque solution est représentée par un chromosome qui est une version codée de toutes ses caractéristiques sous la forme d'une chaîne de symbole
- d'une fonction f d'évaluation des solutions ; f n'est pas forcément une fonction mathématique de chromosomes. Elle peut aussi consister à une simulation de la performance des solutions dans le problème dans lequel elle doit s'appliquer ; cela permet de modéliser le principe de sélection naturelle.

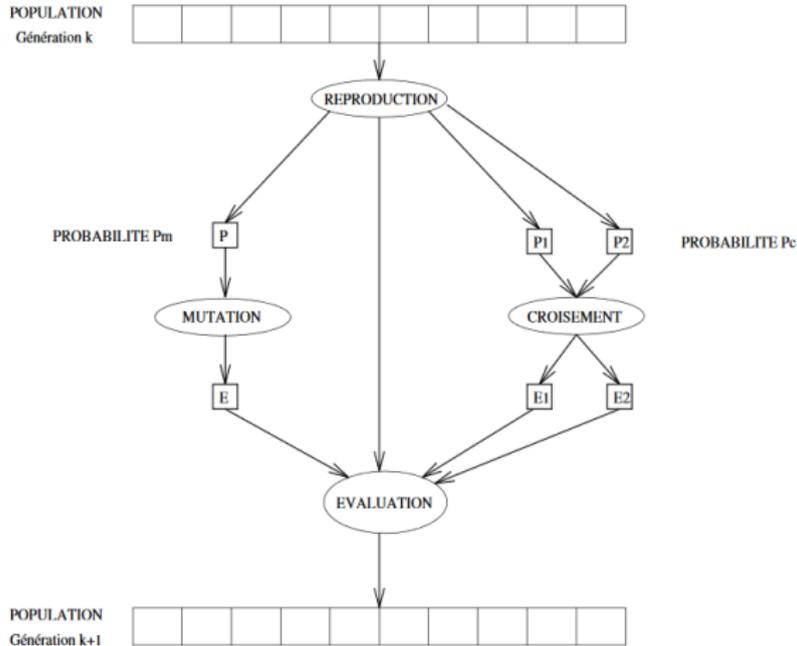
Différents types de méthodes

Algorithmes génétiques

- A chaque itération, l'algorithme génère une nouvelle génération de solutions en appliquant des opérateurs de mutations et de croisements sur des membres de la population courante.
- Une nouvelle population est ainsi constituée à partir de la population courante en éliminant certaines solutions peu performantes et en créant des nouvelles par des mutations et des croisements
- L'application itérative de la procédure conduit à l'optimisation de la population selon les critères d'évaluation choisis au départ

Différents types de méthodes

Algorithmes génétiques



Trois types de données

Le point commun du machine learning et du deep learning est d'exploiter des données pour l'entraînement de modèles probabilistes. Avec les algorithmes et/ou logiciels et le matériel, les données sont la troisième composante clé de la plupart des IA de nos jours.

- Trois types de données
 - ▶ Données d'entraînement
 - ▶ Données de test
 - ▶ Données de production

Données d'entraînement

- Les jeux de données qui vont servir à entraîner un modèle de machine learning ou de deep learning pour en ajuster les paramètres.
- Les données sont étiquetées à l'avance pour faciliter l'apprentissage.
- Plus la base est grande, meilleure sera l'entraînement du système
- Il est évidemment nécessaire de disposer de données d'entraînement de qualité

Données de test

- Les données qui serviront à vérifier la qualité de l'entraînement d'un système
- Les données doivent avoir une distribution statistique voisine des données d'entraînement, au sens où elles doivent être bien représentatives de la diversité des données que l'on trouve dans la base d'entraînement et que l'on aura dans les données de production
- Les données de tests sont un sous-ensemble d'un jeu de départ dont une partie sert à l'entraînement et une autre partie, plus limitée, sert aux tests.
- Elle permettra d'identifier le taux d'erreur du système ; le niveau de taux d'erreur acceptable dépend de l'application

Données de production

- Il s'agit des données non étiquetées qui alimenteront le système lors de son utilisation en production pour faire une prévision des étiquettes manquantes
- Les données d'entraînement sont normalement anonymisées pour l'entraînement du système
- Par contre, les données de production peuvent être nominatives ainsi que les prévisions associées générées par la solution