

UF Analyse descriptive et prédictive

Contexte :

- L'Intelligence Artificielle
- L'apprentissage



M.-J. Huguet
<https://homepages.laas.fr/huguet>
2018-2019



Références (1)

- Intelligence Artificielle, S. Russel, P. Norvig, Pearson Education, 2010
 - Intelligence Artificielle, Livre Blanc INRIA, 2016
 - L'IA, mythes ou réalités, Intersectices 2015
 - https://interstices.info/jcms/p_84122/l-intelligence-artificielle-mythes-et-realites
- Apprentissage Artificiel, A. Cornuejols, L. Miclet, Eyrolles, 3^{ème} Ed., 2018
- Cours (accessibles sur le web):
 - Histoire de l'IA, Frédéric Fürst, Univ Picardie, 2015 (?)
 - Apprentissage, Jean-Daniel Zucker, UPMC, 2010-2011
 - Clustering, Gilles Gasso et Ph. Leray, INSA Rouen
 - Apprentissage, Chloe-Agathe Azencott, Centrale SupElec, 2016-2017
 - Apprentissage non supervisé, Nicolas Baskiotis, Paris6, 2016-2017

Références (2)

- **Cours du collège de France**

- <http://www.college-de-france.fr>
- Enseignement → Mathématiques et Sciences Numériques
 - Gérard Berry : Algorithmes, machines et langages
- Enseignement → Chaires annuelles → Informatique et Sciences Numériques
 - 2017-2018 : Claire Mathieu – Algorithmes
 - 2015-2016 : Yann LeCun : L'apprentissage profond
 - 2011-2012 : Serge Abiteboul : Sciences des données : de la logique du premier ordre à la toile

Références (3)

- Jeux de tests :
 - Opendata : <https://www.data.gouv.fr/fr/>
 - UCI : <http://archive.ics.uci.edu/ml/index.php>
 - Challenge Kaggle : <https://www.kaggle.com/>
- Meetup Toulouse Data Science
 - <https://www.meetup.com/fr-FR/Tlse-Data-Science/>

Références (4)

- **Associations scientifiques :**

- Association Française d'IA : <http://afia.asso.fr/>
- Groupe de Recherche IA (CNRS) : <http://www.gdria.fr/>

- **Site (Informatique) :**

- <https://interstices.info/>

Section 1. Contexte : l'Intelligence Artificielle

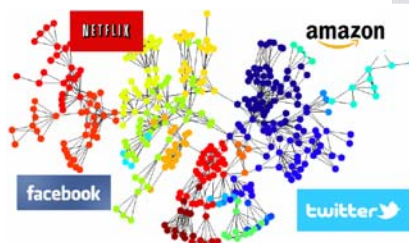
Le buzz (1)

- L'IA est en vogue dans les médias et la presse scientifique
- Nombreuses réalisations depuis 20 ans avec une forte accélération :
 - 1997 : Deep Blue (IBM) : un système informatique bat le champion du monde d'échecs
 - 2005 : Challenge DARPA : un robot se déplace de façon autonome en env. inconnu
 - 2007 : Challenge DARPA : un robot se déplace de façon autonome en env. urbain
 - 2011 : Watson (IM) : bat des champions du monde au Jeu Jeopardy!
 - 2012 : Succès de l'apprentissage profond en reconnaissance d'images (Imagenet)
 - Large Scale Visualization Challenge
 - 2014 : Google : description automatique d'images
 - 2015 : Facebook : reconnaissance faciale
 - 2015/2016 : AlphaGo (Google) : bat le champion du monde de Go
 - 2016 : Daddy's Car (composition style Beatles), Sony CSL
- **Impact des données massives sur le (re)-nouveau de l'IA**
- **Apprentissage « artificiel » ou apprentissage « machine »**

7

Le buzz (2)

- Mais aussi : 2017 ... 2018 ... 2019 ...
 - Compagnons
 - Traduction instantanée
 - Recommandation, identification de communautés
 - Véhicules sans chauffeur



8

Le buzz (3)

- De nombreux outils pour l'apprentissage :
 - Boîte à outils :
 - Scikitlearn (Python), R, Matlab,
 - Deep Learning / Big Data :
 - TensorFlow, Theano, Caffe, Torch,
 - Langages :
 - Julia, Scala, ...
 - Quelques liens :
 - <https://julialang.org/>
 - <https://scala-lang.org/>
 - <https://www.coursera.org/learn/scala-spark-big-data>

Débats (1)

• Débats génériques

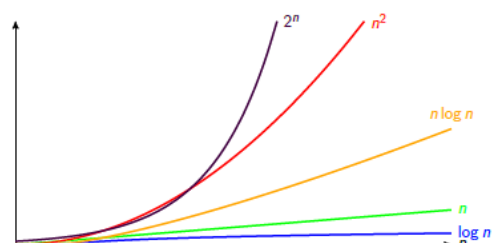
- Intelligence ?
 - logico-mathématique
 - musicale-rythmique
 - verbo-linguistique
 - sociale
 - corporelle-kinesthésique
 - spatiale et naturaliste

• Limites humaines et limites des machines

- **Décidabilité** : Peut-on tout calculer ?
- NON (Turing, 1936)



- **Complexité** : temps de réponse d'un algorithme ?



Débats (2)

- **Débats : intérêt de l'IA ?**

- Robots guerriers
- Transhumanisme
- Emploi
- Trading haute fréquence
- Fouille de données et respect de la vie privée
- Coût écologique
- ...
- Crainte du pouvoir « donné » aux machines :
 - Éthique des recherches et des algorithmes,
 - Transparence, explicabilité
 - Autonomie partagée

11

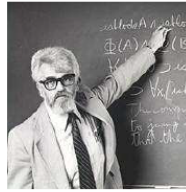
Le buzz (5)

- Quelques thèmes de l'Intelligence Artificielle
 - Vision, parole
 - Ingénierie des connaissances, web sémantique
 - Robotique, véhicules autonomes
 - Traitement du langage naturel
 - Raisonnement et aide à la décision
 - ...
- Apprentissage : problème central en IA

12

Bref aperçu de l'IA (1)

- 1956 : Naissance « officielle » de l'IA
 - Conférence de Dartmouth (initiateurs :)



J. McCarthy
(Dartmouth College)



M. L. Minsky
(Harvard University)



N. Rochester
(I.B.M. Corporation)



C.E. Shannon
(Bell Telephone Laboratories)

Rêve de l'IA : Construire des machines pouvant égaler l'Homme

- *Comment simuler la pensée et le langage au travers de règles formelles?*
- *Comment faire penser un réseau de neurones?*
- *Comment doter une machine de capacité d'apprentissage automatique?*
- *Comment doter une machine de créativité?*

13

Bref aperçu de l'IA (2)

- Précurseurs (années 1950 ...) :
 - A. Turing : une machine peut-elle penser ? (Test de Turing)
 - H. Simon & A. Newel : démonstration automatique de théorèmes (Logic Theorist)
 - W. McCulloch & W. Pitts : réseaux de neurones artificiels
 - W. Weaver : règles linguistiques, traduction automatique
- 1956 – 1974 : l'Age d'or
 - 1957 : **Perceptron** (F. Rosenblatt) : réseau de neurones artificiels à 1 seule couche
 - 1957 : **GPS General Problem Solver** (H. Simon & A. Newel)
 - Problème = Etat Courant, But, Opérateurs
 - Systèmes de production = ensemble de règles d'inférence
 - Résolution : appliquer ces règles (ch. arrière / induction)
 - 1958 : **LISP** (J. McCarthy) : LISP Processing, langage de référence de l'IA
 - Représentation du programme et des données par des listes
 - Année 60 : **Grammaires** de N. Chomsky, Réseaux Sémantiques (R. Quillian)

14

Bref aperçu de l'IA (3)

- 1956 – 1974 : l'Age d'or
 - Problèmes de jeux, Démonstration, Heuristiques, Logique, Planification d'actions, Traitement du langage naturel, Vision, ...
 - Deux familles :
 - Logique et informatique
 - Sciences cognitives et sciences humaines
 - **Des prédictions (1958) ... dans 10 ans ...**
 - un ordinateur sera champion du monde d'échecs
 - 1997 : Deep Blue
 - un ordinateur écrira de la musique qui sera considérée comme esthétique
 - 2016 : Daddy's Car (style des Beatles),
 - un ordinateur découvrira un théorème fondamental des mathématiques
 - les théories psychologiques auront la forme de programmes

15

Bref aperçu de l'IA (4)

- 1974 – 1980 : Premier hiver de l'IA
 - Crise scientifique, morale et financière
 - Limite en puissance de calcul, Limite NP-Complétude
 - Besoin de grandes masses de données pour produire des connaissances
 - Paradoxe :
 - certains problèmes difficiles pour l'Homme peuvent être résolus (ex : démonstration) mais des problèmes faciles pour l'Homme (reconnaître un visage), pas de résultats probant
 - Est-il pertinent de doter les machines d'intelligence ?
 - L'intelligence peut-elle se ramener à des modèles symboliques ?
 - Développement de nouveaux modèles et méthodes :
 - Extension des logiques : Multi-valuées, Floues, Modales, Temporelles, ...
 - Langage Prolog (clauses de Horn),

16

Bref aperçu de l'IA (5)

- 1980-1987 : le retour des financements ...
 - Boom des systèmes experts et de l'ingénierie des connaissances (E. Feigenbaum)
 - Application à des domaines spécifiques (demande des entreprises)
 - Pour limiter l'explosion combinatoire des systèmes de production
 - Premiers systèmes experts : DENDRAL (Chimie), MYCIN (Médical)
 - Nouveaux modèles d'acquisition et de représentation des connaissances
 - Graphes conceptuels, Logique de description
 - Raisonnement et Décision :
 - Programmation par contraintes, Raisonnement à base de cas,
 - Apprentissage symbolique
 - Permettre aux programmes de s'améliorer
 - Liens avec la biologique
 - Algorithmes génétiques, Insectes sociaux, Systèmes Multi-Agents
 - Réseaux de neurones : rétro-propagation du gradient (J. Hopfield)
 - Applications à la reconnaissance de caractères ou reconnaissance vocale

17

Bref aperçu de l'IA (6)

- 1987 – 1993 (1995 ?) : Second hiver de l'IA
 - Spécialisation/Fragmentation de l'IA en nombreux sous-domaines
 - Recentrage vers la robotique
 - Le corps (capacités sensorielles, mouvement) est essentiel pour produire de la connaissance et du raisonnement
 - J. Pearl : liens entre des problèmes d'IA et d'autres domaines (mathématique, économie, recherche opérationnelle) : probabilité, théorie de la décision, optimisation, modèles de Markov, réseaux bayésiens, ...
 - Nombreuses réalisations dans des applications spécifiques :
 - Robotique industrielle, reconnaissance vocale, diagnostic, reconnaissance des formes, recherche d'information, ...
- Quand un problème est résolu, ce n'est plus un problème d'IA ...

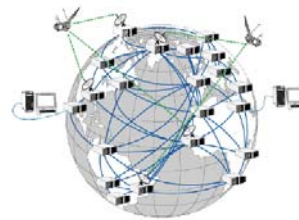
18

Bref aperçu de l'IA (7)

- Fin années 1990 : Retour en force de l'IA

- Années 1990 :

- développement des ordinateurs personnels et du Web
 - Rendre les moteurs de recherche plus intelligent
 - Web sémantique : des données aux connaissances / Ontologies
 - Jeux vidéo



- Pionniers (en France)

- Jacques Pitrat (Doct. 1966) et Jean-Louis Laurière (1976 – ALICE) - Paris 6
 - Alain Colmerauer (Doct. 1967, - Prolog1 1974, Prolog4 1996) - Marseille

Bref aperçu de l'IA (8)

- **Et maintenant :**

- Données massives, Apprentissage Automatique
 - Convergence de différentes évolutions
 - Apport intelligence computationnelle et puissance de calcul
 - Avancées algorithmiques
 - Améliorations méthodologiques, calcul haute performance
 - Capacité de calcul et de stockage
 - Processeurs, architecture,
 - Explosion des données
 - Des chiffres, des chiffres, ...

Section 2. L'apprentissage automatique

21

Définition générale

- **Qu'est-ce que l'apprentissage pour une « machine » ?**
 - Simon (1983) : *changements dans un système lui permettant d'améliorer ses performances sur une tâche déjà vue ou similaire.*
 - **Tâches ? Applications ?**
 - Prédiction ? Classification ? Planification et Action ? Résolution de problèmes ?
 - **Changements ?**
 - Acquérir de nouvelles connaissances / de nouvelles capacités
 - Réviser ses connaissances / son comportement
 - Résoudre mieux / plus efficacement les problèmes
 - **Performances ?**
 - Du processus d'apprentissage,
 - Des capacités du système après apprentissage, ...
 - **Apprentissage machine / apprentissage artificiel**

22

Exemple

- **Apprendre à jouer aux échecs**

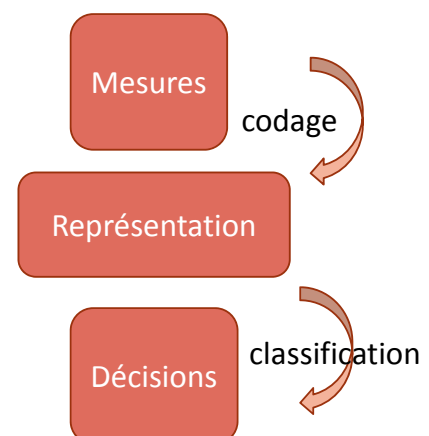
- But :
 - Apprendre les règles du jeu par observation ? **Apprendre à bien jouer ?**
 - Prédire les actions de l'adversaire ? Prédire le vainqueur ? ...
- Méthode de résolution pour chercher le prochain coup à jouer
 - algo MinMax basé sur une fonction d'évaluation
- Apprendre une « bonne » fonction d'évaluation
 - Identifier des stratégies par rapport à des positions, Apprendre par cœur certains coups, ...
- Choix des données d'apprentissage :
 - quelles parties d'échecs ? Intérêt de prendre en compte des coups « faibles », des coups illégaux ? Quel séquençement ?
- Comment valider après apprentissage :
 - nombre de parties gagnées (sur quel échantillon ?) avec explication ?

23

Exemple

- **Reconnaître des caractères manuscrits**

- Quels caractères : séparation des éléments ?
- Difficulté : impossible d'apprendre par cœur toutes les formes d'écriture
 - Codage des données : matrice binaire
 - Pré-traitement des images et segmentation
 - Homogénéisation (centrage, échelle,)
 - Reconnaissance (classification)
 - Descripteurs pertinents
 - Hauteur, largeur, boucle, ...
 - Post-traitement : spécificités linguistique



- **Reconnaissance des Formes** : Théorie du Codage (de l'information), Traitement du signal, Traitement d'images, ...

24

Raisonnement

- **Différents mécanismes de raisonnement**

- **Raisonnement inductif** : de l'expérimentation à la production de connaissances

- Déterminer des « lois » générales à partir de cas particuliers
 - Partir d'un nombre fini de propositions pour en proposer une conclusion
- Supervisé (avec professeur) / non supervisé (sans professeur)
 - Ex : classification, partitionnement, identification, ...

- **Raisonnement déductif** : de l'explication à la révision de connaissances

- Déterminer une conclusion particulière à partir d'affirmations générales
 - La conclusion est une conséquence nécessaire des affirmations par inférence logique
- Environnement connu / inconnu
 - Ex : Résolution de problèmes, action

25

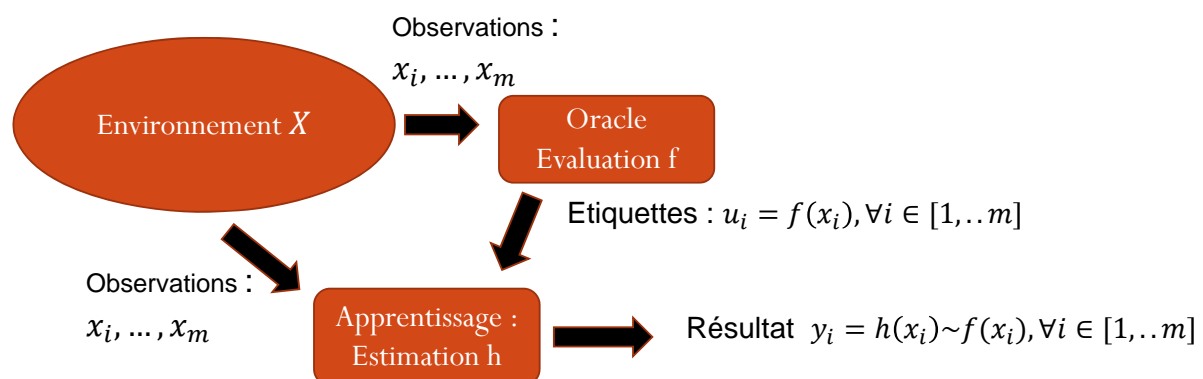
Types d'apprentissage (1)

- **Différents types d'apprentissage**

- **Apprentissage supervisé**

- Deux étapes

- 1. Déterminer un modèle à partir de données étiquetées



26

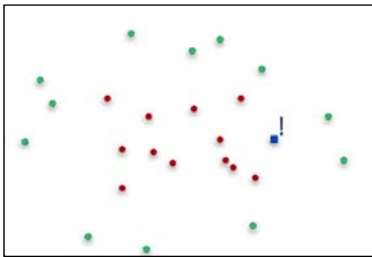
Types d'apprentissage (2)

- **Différents types d'apprentissage**

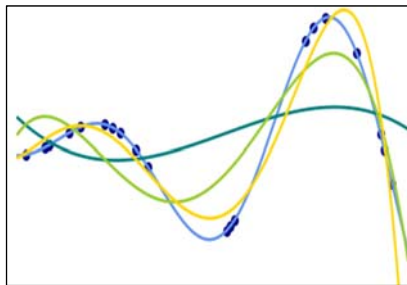
- **Apprentissage supervisé**

- Deux étapes

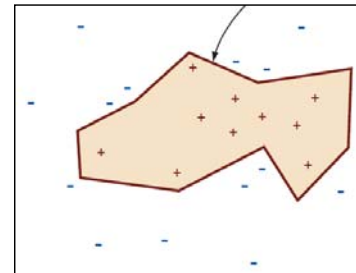
- 2. Déterminer l'étiquette d'une nouvelle donnée, connaissant le modèle appris



Classification
(y : fonction discrète ou classe)



Régression
(y : fonction continue)



Concept
(y : fonction booléenne)

Types d'apprentissage (3)

- **Différents types d'apprentissage**

- **Apprentissage non supervisé**

- Le système ne dispose que d'exemples :

- Données X sans étiquette

- Nombre et nature des classes inconnu

- Rechercher une structure dans les données

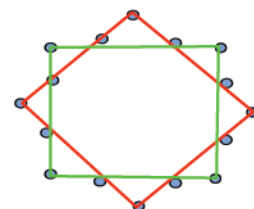
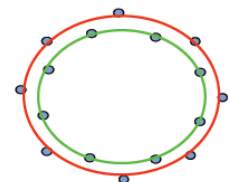
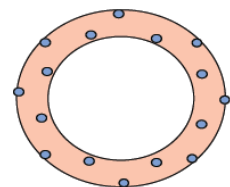
- Partitionner les exemples en clusters/classes

- Clustering (segmentation, partitionnement)

- Partitionner les exemples en clusters/classes

- **Un bon clustering ?**

- Homogènes : les éléments d'un même cluster sont similaires
 - Séparés : les éléments de différents clusters sont différents



Types d'apprentissage (4)

- **Différents types d'apprentissage**

- **Apprentissage par renforcement**

- Apprendre par interaction avec l'environnement, par rapport à des observations
 - Observation d'un état
 - Calcul d'une action (prise de décision)
 - Réalisation de l'action
 - Réception (et mémorisation) d'un résultat suite à cet action
- Découvrir une stratégie via un mécanisme de récompenses (positives ou négatives)

Apprentissage Hors-Ligne / En-Ligne

- **Protocole d'apprentissage**

- Caractérisation des interactions entre apprenant et environnement
 - Disponibilité/actualisation des données

- **Apprentissage hors-ligne (batch)**

- Les données sont disponibles et stables

- **Apprentissage en-ligne**

- Traitement « à la volée » et incrémental

Apprentissage descriptif/prédictif/descriptif

- **Apprentissage descriptif**

- Comprendre les données; les décrire; chercher régularité
 - Apprentissage non-supervisé
 - Ex : pourquoi y-a-t-il des bouchons sur la route ?

- **Apprentissage prédictif**

- Identifier des règles de décision / données futures
 - Apprentissage supervisé
 - Ex : quel sera le trafic dans 2 heures ?
- Ex : diagnostic médical - Description des patients – Prédiction d'une pathologie

- **Apprentissage prescriptif**

- Quelles actions effectuer ?
 - Ex : quel est le meilleur trajet en partant à 8h ?

31

Apprentissage Automatique (1)

- **Science pour la modélisation**

- Mettre en évidence des propriétés/régularités sur un ensemble d'observations
- Déterminer un modèle d'un environnement/système permettant la prédiction ou la prise de décision

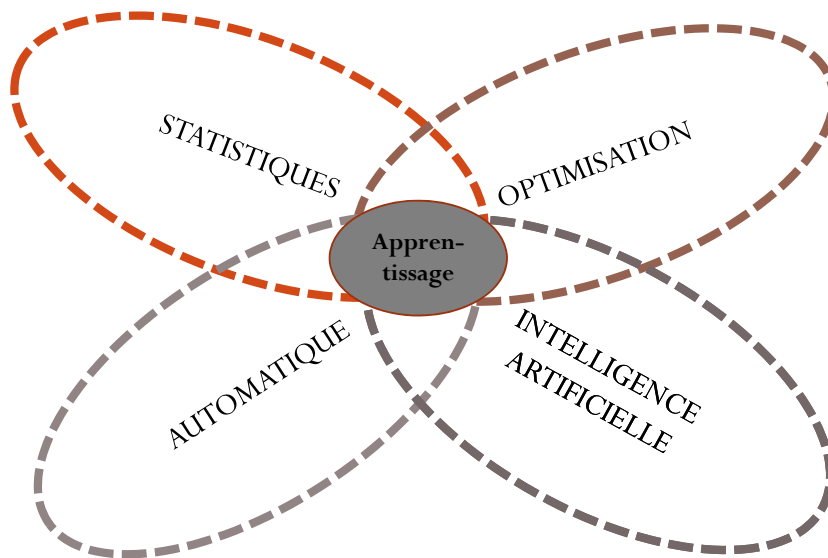
- **Science des systèmes adaptatifs**

- Evolution par simulation
- Apprentissage par Renforcement pour l'action

32

Apprentissage Automatique (2)

- **Domaine pluridisciplinaire**



33

Plan de l'UF

- **Partie Analyse exploratoire – G.Trédan – M. Roy – R. Pasqua**

- Méthodologie pour analyser un ensemble de données
- Visualisation de données
- TP en R

- **Partie Apprentissage supervisé – M-V. Le Lann – M. Siala**

- Principes généraux
- Quelques méthodes : K-plus proches voisins; Réseaux de neurones, Random Forest, Support Vector Machines
- TP en Python (ScikitLearn)

- **Partie Apprentissage non-supervisé – MJ. Huguet – M. Siala**

- Principes généraux
- Quelques méthodes : k-means, approches hiérarchiques, basées graphes
- Quelques problèmes : fouille de données, réseaux sociaux
- TP en Python

34