

# Apprentissage Artificiel

Engelbert Mephu Nguifo  
Université Clermont Auvergne

## Informations

- CM: E. Mephu Nguifo (8 h)
- TD : A. Plaud et V. Siyou (8 h)
- TP : A. Plaud et V. Siyou (8 h)
  
- Evaluation : Examen, DS, et contrôle TP

## Syllabus

- 1 : Introduction : de l'apprentissage naturel à l'apprentissage artificiel  
Processus algorithmique, Thèse de Church, Test de Turing, Chambre de Searle...
  - 2 : Machine de Turing  
Définition, Principe, Modèle de Turing, Modèle de calcul;
  - 3 : Environnement méthodologique de l'apprentissage  
Espace des données, espace des hypothèses, protocoles, évaluation, comparaison
  - 4 : Apprentissage par exploration  
Espace des versions, Apprentissage par analogie, Raisonnement à partir de cas, ...;
  - 5 : Apprentissage par optimisation  
Arbres de décisions, réseaux connexionnistes, ...
- Mise en œuvre avec un outil logiciel (WEKA, SAS, ...)

## Objectifs

- Introduction à un thème de recherche du laboratoire LIMOS : Apprentissage et fouille de données
- Compétences :
  - Connaître les principes de base de l'apprentissage artificiel
  - Connaître des techniques d'apprentissage supervisé
  - Savoir utiliser des logiciels d'apprentissage supervisé

## Bibliographie

- Apprentissage Artificiel, 2010, chez Eyrolles, par A. Cornuéjols et L. Miclet
- Data Mining: concepts and techniques, 2012, par J. Han, M. Kamber et J. Pei
- Machine Learning, 1997, par T. Mitchell, MIT Press
- Encyclopedia of Machine Learning, 2012, par C. Sammut et G. I. Webb

## Webographie

- [www.kdnuggets.com](http://www.kdnuggets.com)

## Introduction

- Apprentissage ?
  - Sciences cognitives:
    - Capacité à améliorer les performances au fur et à mesure de l'exercice d'une activité
      - Joueur des échecs, plus généralement tout joueur
      - Enfant qui apprend
    - + Faculté à généraliser
  - Larousse:
    - Ensemble des processus de mémorisation mis en œuvre par l'animal ou l'homme pour élaborer ou modifier les schèmes comportementaux spécifiques sous l'influence de son environnement et de son expérience.

## Introduction

- Apprentissage ?
  - Cas de l'enfant qui apprend :
    - Apprentissage par cœur
    - Par instruction
    - Par généralisation
    - Par découverte
    - Avec des catégorisations, voire formation de théories
    - Plus ou moins supervisé ou autonome
    - Etc ...

## Introduction

- Apprentissage ?
  - Intelligence Artificielle:
    - Alan Turing
      - Computing Machine and Intelligence, Revue Mind, octobre 1950 (vol. LIX, numéro 236)
        - » Paragraphe : « Learning Machines »
        - » <http://data.archives.ecs.soton.ac.uk/turing>
        - » <http://www.abelard.org/turpap/turpap.htm>

## Introduction

- Apprentissage Artificiel ?
  - Tom Mitchell, chercheur au Carnegie Mellon (US):
    - écriture de programmes d'ordinateur capables de s'améliorer automatiquement au fil du temps, soit sur la base de leur propre expérience, soit à partir de données antérieures fournies par d'autres programmes
  - Dictionnaire

## Introduction

- Apprentissage Artificiel ?
  - Wikipedia.fr (29/01/2013)
    - discipline scientifique concernée par le développement, l'analyse et l'implémentation de méthodes automatisables qui permettent à une machine (au sens large) d'évoluer grâce à un processus d'apprentissage, et ainsi de remplir des tâches qu'il est difficile ou impossible de remplir par des moyens algorithmiques plus classiques

## Introduction

- Apprentissage Artificiel ?
  - Rôle essentiel :
    - Reconnaissance de la parole
    - Vision par ordinateur
    - Traitement automatique de texte
    - Robotique
    - **Reconnaissance de formes**
  - Rôle sans cesse croissant
    - Compréhension de plusieurs principes calculatoires qui guident tout processus d'apprentissage (humain ou machine)

## Introduction

- Apprentissage Artificiel ?
  - Riches fondements théoriques
    - Combien au minimum d'exemples d'entraînement faut-il fournir à un programme d'apprentissage pour être certain qu'il apprenne avec une efficacité donnée ?
      - Théorie statistique de l'estimation
      - Théorie de l'identification et de la commande optimale
      - Complexité de l'apprentissage de grammaires
      - Inférence bayésienne algorithmique

## Introduction

- Apprentissage Artificiel ?
  - Applications
    - Aide à la décision des experts dans des environnements complexes et évolutifs
      - Marché financier
      - Diagnostic médical
      - Marketing (ciblage de clientèle pour les campagnes de publicité)
      - Attrition (Churn en anglais): abonnés fuyants
    - Analyse automatique de photos satellites pour détecter certaines ressources sur la Terre
    - Jeux
      - 11 mai 1997 : Echec de Gary Kasparov contre un programme de jeux d'échecs
      - RoboCup : [www.robocup.org](http://www.robocup.org), compétition annuelle robot footballeur

## Introduction

- Apprentissage Artificiel ?
  - Programmer les ordinateurs pour leur faire faire exécuter des tâches considérées comme intelligentes, de multiples façons et de manière efficace
  - Discipline jeune, dont les outils et champs d'applications sont disparates
  - Nécessité d'avoir une culture généraliste:
    - Algèbre linéaire, Probabilités, Combinatoire,
    - Logique, Analyse élémentaire, Algorithmique
    - Théorie des langages

## Introduction

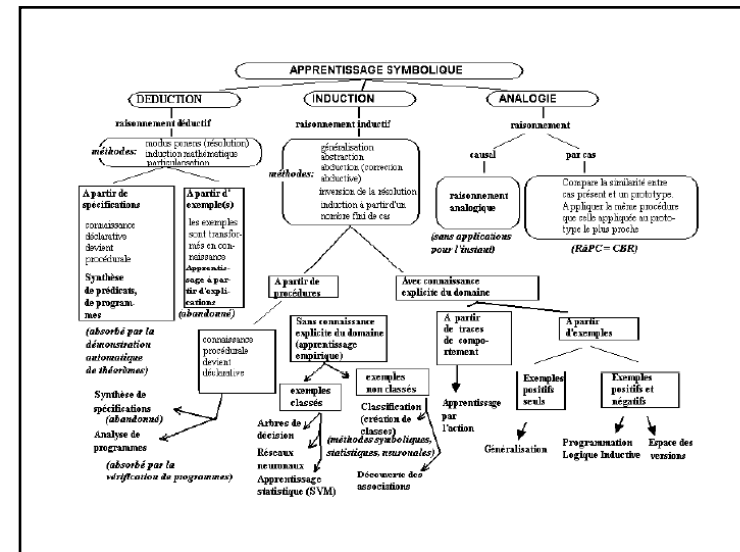
- Apprentissage Artificiel ?
  - Apprentissage automatique, ou en anglais: Machine Learning
  - Méthode permettant de construire un modèle de la réalité à partir de données, soit en améliorant un modèle partiel ou moins général, soit en créant complètement le modèle. (Yves Kodratoff)
  - Deux tendances:
    - Apprentissage symbolique
    - Apprentissage numérique ou statistique

## Introduction

### • Apprentissage Artificiel ?

#### – Termes

- Fouille de données ou Extraction de connaissances à partir des données
- Précision versus Généralisation : dilemme de l'apprentissage
  - Précision : écart entre une valeur mesurée et une valeur réelle
  - Sur-Apprentissage : apprentissage avec trop de précision
  - Sur-généralisation : apprentissage avec trop peu de précision
- Intelligibilité (ou Comprehensibility or Understandability)
- Classification, classement, regression
  - Classification : en anglais, clustering
  - Classement : en anglais, classification
  - Régression

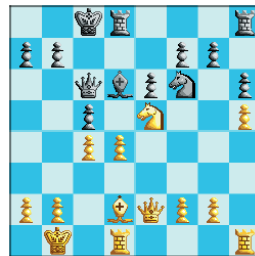


## Introduction

### • Apprentissage Artificiel ?

#### – Exemple : Apprendre à jouer

- Jeux à 2 adversaires, sans hasard, sans information cachée
  - Jeux d'échec,
  - Jeu de Go
  - Morpion



## Introduction

### • Apprentissage Artificiel ?

#### – Exemple : Apprendre à jouer

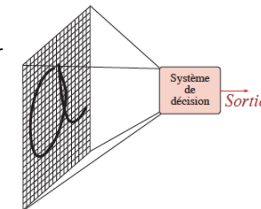
- But poursuivi :
  - Apprendre les règles du jeu à partir d'observations
  - Apprendre à bien jouer
  - Faire découvrir les variables pertinentes pour prédire comment jouera l'adversaire (niveau, style de jeu)
  - Prédire le gagnant ou le nombre de coups restant à jouer
  - ....
- Situation de jeu :
  - Choix entre plusieurs dizaines de coups légaux: lequel est le meilleur
  - Comment déterminer le coup à jouer ? (variables : durée et nombre de coups restant, ...)
  - Jouer le coup permettant de maximiser un gain
- Apprendre = déterminer la fonction d'évaluation qui donne la qualité des solutions

## Introduction

- Apprentissage Artificiel ?
  - Exemple : Apprendre à jouer
    - Autres points
      - Apprendre par cœur
      - Choix des données d'apprentissage
      - Séquencement des leçons
      - Validation de l'apprentissage

## Introduction

- Apprentissage Artificiel ?
  - Exemple : Apprendre à lire
    - Reconnaissance de caractères manuscrits sur enveloppe timbrée
      - Codage
        - » Variété des formes : infini
        - » Pas d'apprentissage par cœur
        - » Généralisation

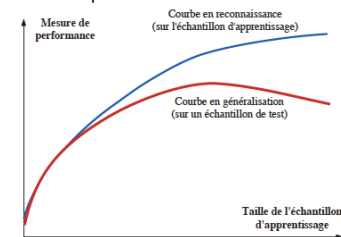


## Introduction

- Apprentissage Artificiel ?
  - Exemple : Apprendre à lire
    - Reconnaissance de caractères manuscrits sur enveloppe timbrée
      - Mesure de performance
        - » Nombre d'erreurs (avec proba mauvaise classification) ?
        - » Mesure de risque plus élaborée ?
          - Erreur sur une lettre <> Erreur sur un chiffre

## Introduction

- Apprentissage Artificiel ?
  - Exemple : Apprendre à lire
    - Reconnaissance de caractères manuscrits sur enveloppe timbrée
      - Mesure de performance



## Introduction

- Apprentissage Artificiel ?
  - Exemple : Apprendre à lire
    - Reconnaissance de caractères manuscrits sur enveloppe timbrée
      - Modélisation
        - » Forme de l'apprentissage à réaliser
          - Approche géométrique : apprendre des frontières
          - Approche probabiliste
          - Critère de décision par les plus proches voisins
            - Nécessité d'un nombre suffisant de points étiquetés et d'une relation de voisinage

## Introduction

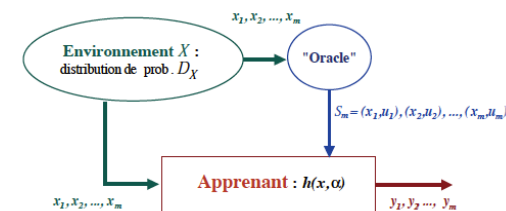
- Apprentissage Artificiel ?
  - Evolution, Transformation, Modification
  - Lien
    - Cybernétique
    - Sciences cognitives

## Introduction

- Apprentissage Artificiel ?
  - Concepts de base
    - Cadre de l'induction supervisé
      - Processus par lequel on tire des lois de portée générale en partant de l'observation de cas particuliers
      - Nécessité de disposer d'un expert (oracle) qui donne les étiquettes de classe
      - Données : couple (observation, sortie désirée ou étiquette)
 
$$z_i = (x_i, u_i)$$

## Introduction

- Apprentissage Artificiel ?
  - Concepts de base
    - Cadre de l'induction supervisé



## Introduction

- Apprentissage Artificiel ?
  - Concepts de base
    - Notions clés
      - Critère de succès
        - » Approximer au mieux la sortie désirée pour chaque observation
        - » Fonction de risque, évaluer la performance
      - » Simplicité du résultat d'apprentissage produit
      - » Compréhensibilité, intelligibilité
      - » Facilité d'intégration dans la théorie courante
      - » Faible coût computationnel

## Introduction

- Apprentissage Artificiel ?
  - Concepts de base
    - Notions clés
      - Protocole d'apprentissage
        - » Apprentissage hors ligne (batch learning)
          - Toutes les données sont fournies d'un seul coup
        - » Apprentissage séquentiel, incrémental ou en ligne (on-line learning)
          - Données en séquences
          - Réponse après chaque entrée ou groupe d'entrées
      - » Tâches : identification versus prédiction

## Introduction

- Apprentissage Artificiel ?
  - Concepts de base
    - Notions clés
      - Protocole d'apprentissage
        - » Apprenant +/- actif
          - Passif
        - Ou initiative limitée dans la recherche d'informations
          - Orienter la distribution de probabilités
          - Poser des questions sur les classes d'une observation
          - Mieux organiser les expériences

## Introduction

- Apprentissage Artificiel ?
  - Concepts de base
    - Notions clés
      - Tâche d'apprentissage
        - » Connaissance
          - Modifier le contenu
          - Acquisition de connaissance, révision
        - » Environnement
          - Reconnaître des formes: supervisé / non supervisé
          - Prédire
          - Etre plus efficace

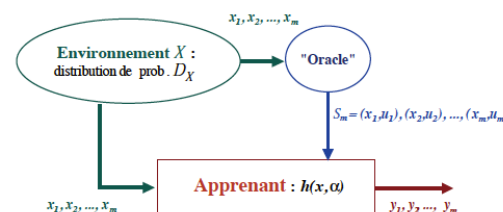


## Introduction

- Apprentissage Artificiel ?
  - Concepts de base
    - Notions clés
      - Tâche d'apprentissage
        - » Classes abstraites de problèmes
          - Caractériser l'apprentissage
            - Compression d'information, Cryptographie, ....
        - » Structures de données ou types d'hypothèses
          - Langage d'expression des hypothèses
            - Expression booléenne, grammaire, fonction linéaire, programme logique, réseaux bayésiens,

## Introduction

- Apprentissage Artificiel ?
  - Concepts de base
    - Estimation de fonction
      - Ensemble d'exemples  $S = \{x_i, u_i\}_{1 \leq i \leq m} = \{x_i, f(x_i)\}_{1 \leq i \leq m}$



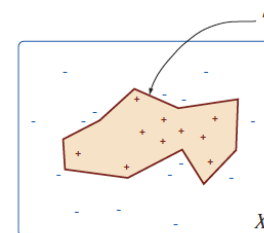
## Introduction

- Apprentissage Artificiel ?
  - Concepts de base
    - Estimation de fonction
      - Problème de régression : problème d'estimation
        - » Trouver une fonction  $h$  qui approxime  $f$ 

$$\forall x \in \mathcal{X}, h(x) \approx f(x) = u$$
    - Apprentissage d'une classification d'objets
      - » Valeurs discrètes
    - Apprentissage de concept
      - » Reconnaître une classe d'objets parmi tous les objets
    - Optimisation multicritères

## Introduction

- Apprentissage Artificiel ?
  - Un Jeu entre espaces
    - Apprentissage de concepts

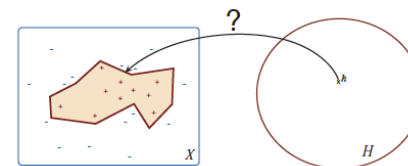


## Introduction

- Apprentissage Artificiel ?
  - Un Jeu entre espaces
    - Apprentissage de concepts
      - Apprentissage est impossible ...
  - ... sans limiter l'espace des hypothèses

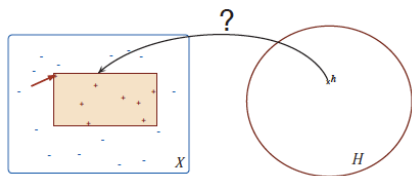
## Introduction

- Apprentissage Artificiel ?
  - Un Jeu entre espaces
    - Apprentissage de concepts



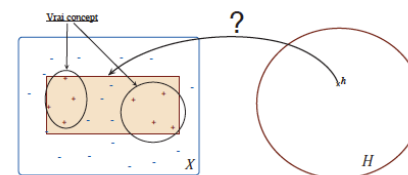
## Introduction

- Apprentissage Artificiel ?
  - Un Jeu entre espaces
    - Apprentissage de concepts



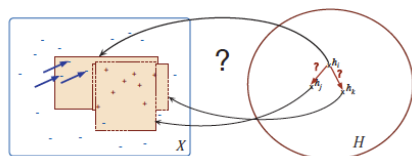
## Introduction

- Apprentissage Artificiel ?
  - Un Jeu entre espaces
    - Apprentissage de concepts



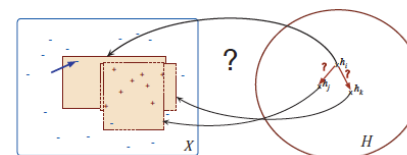
## Introduction

- Apprentissage Artificiel ?
  - Un Jeu entre espaces
    - Apprentissage de concepts
      - Exploration de l'espace des hypothèses



## Introduction

- Apprentissage Artificiel ?
  - Un Jeu entre espaces
    - Apprentissage de concepts
      - Exploration de l'espace des hypothèses



## Introduction

- Apprentissage Artificiel ?
  - Histoire

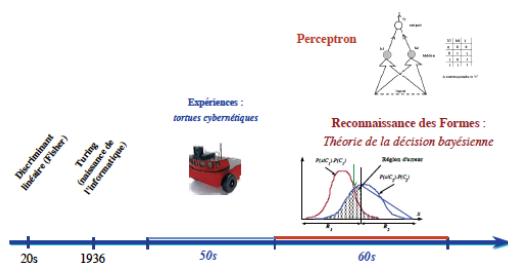


FIG. 1.13 La première période de l'apprentissage artificiel.

## Introduction

- Apprentissage Artificiel ?
  - Histoire

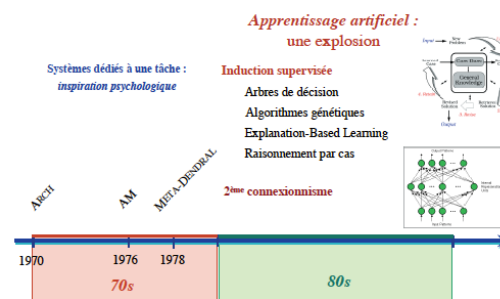


FIG. 1.14 La deuxième période de l'apprentissage artificiel.

## Introduction

- Apprentissage Artificiel ?
  - Histoire

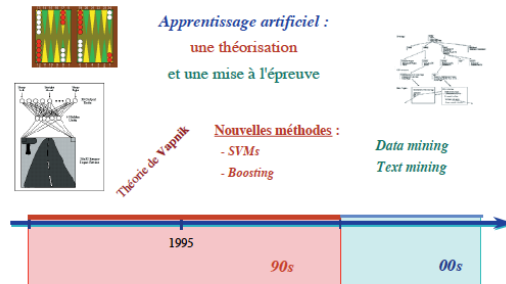


FIG. 1.15 – La troisième période de l'apprentissage artificiel.

## Introduction

- Apprentissage Artificiel ?
  - 12 Conseils Société Oracle
    - Extraire à partir d'encore plus de données
    - Créer de nouvelles variables pour mieux faire parler les données
    - Utiliser une stratégie « en surface d'abord »
    - Construire rapidement plusieurs modèles explicatifs
    - Oublier les pratiques traditionnelles d'hygiène en matière de données
    - Enrichir les données de données extérieures

## Introduction

- Apprentissage Artificiel ?
  - 12 Conseils Société Oracle (suite et fin)
    - Segmenter d'abord les clients et construire des modèles multibuts
    - Construire automatiquement les modèles
    - Varier les sources de données
    - Interpréter les résultats en termes du domaine d'application par une méthode de rétro-ingénierie
    - Compléter les données manquantes
    - Utiliser plusieurs modèles de prédiction à la fois

