

Intelligence artificielle

Exemples d'utilisation et tentative de définition

Gilles Stoltz



Directeur de recherche CNRS, au Laboratoire de mathématiques de l'U. Paris-Sud
Membre du bureau de l'Agence maths-entreprises, pour les Pays de la Loire
Professeur affilié à HEC Paris

... et Lavallois depuis cet été (vive le TGV!)

Détection de spams

Objectif – Classifier chaque courriel comme spam ou non

Données – Une base de données de courriels déjà étiquetés

Méthode – [...]

C'est pourquoi on veut un retour des utilisateurs !

Attribution de crédits

Objectif – Classifier chaque client comme éligible ou non à tel crédit

Données – Une base de données de crédits déjà accordés, en cours de remboursement

Méthode – [...]

La banque étiquette au fur et à mesure les clients

lors d'un défaut, ou à la fin du crédit

Versus : la bonne vieille « règle du 1/3 »

Facile à expliquer mais moins puissante

Diagnostics médicaux

Objectif – Détecter la présence d'une tumeur sur une radiographie

Données – Une base de données de radiographies déjà étiquetées par des radiologues

Méthode – [...]

Difficulté / coût à construire la base de données

Utilisation : complément à l'expertise humaine / déserts médicaux

Détection de fraudeurs par le fisc

Objectif – Classer chaque contribuable comme potentiellement fraudeur ou non

Données – Base de données des revenus, composition familiale, etc. + informations publiques collectées sur les réseaux sociaux ou sur le web (textes, images, vidéos)

Méthode – [...]

Difficulté à traiter de manière automatisée les données hétérogènes
Et il faut le faire pour un grand nombre de contribuables !

C'est un travail que l'homme ne pourrait pas matériellement faire
La machine ne vole pas le travail de l'homme !

Autres exemples de problèmes de **classification**

- Reconnaissance d'images / Détection de caractères
→ Ex : **radars routiers**
- Défense : détection d'appareils étrangers

Importance de la base de données d'entraînement des algorithmes

Apprentissage par exemples étiquetés par des **humains**, pas ex nihilo

Problèmes de **prévision** (= un second objectif possible)

Prévisions de volumes de vente pour l'e-commerce

Travail impossible pour l'homme : trop de produits, trop de variabilités

Décisions de justice

(dommages-intérêts, montants d'une pension, etc.)

Pricing pour chantiers de construction ?

(Règles déterministes + ajustements à partir d'une base de données)

Sources de données nouvelles — les objets connectés

Capteurs connectés pour animaux d'élevage

→ détection précoce de maladies

Compteurs Linky

→ mieux prévoir (voire piloter) la consommation et donc la production d'électricité

Smart homes, smart buildings, etc.

→ difficulté de la prévision / exploitation de comportements individuels

Je n'en parle pas...

- Chatbots (agents conversationnels)
- Gestion automatisée de fonds d'investissement
- Recommandations personnalisées d'achats
- Jeu de go (AlphaGo par Deepmind)

Là encore : jeu de données + objectif + méthode

Triplet

- Objectif / Besoin (sauf exception, n'apparaît qu'après un certain cap)
- Données (nécessite une certaine **maturité organisationnelle**)
- Méthode (l'affaire des matheux, codes déjà écrits et gratuits)

Automatiser uniquement ce qui existe déjà en un sens

Cf. jeu de données existant !

Statistique « traditionnelle » suffit si données « faciles »

Petit volume, et pas de vidéo / texte / image

Ex : Préviation de ventes pour des commerces de détail

Qui était frustré de ne pas appliquer de méthodes statistiques ?

Pas un drame, donc, si méthodes IA pas appliquées

Clé du processus : bon jeu de données

On ne fait que **reproduire** l'existant, mais **efficacement**

- « Garbage in, garbage out »
- Problème de la discrimination (« fair learning »)

Grands volumes de données hétérogènes (ex : réseaux sociaux)

Difficile de les exploiter **automatiquement**, de trouver les pépites

Idée d'une méthode : *k*-plus proches voisins pour classification

Principes

- Difficile de prévoir l'individuel, plus facile de prévoir le collectif
- Le comportement collectif n'est pas assez informatif

(Ex : choix entre deux options totalement équivalentes)

Conséquence : se ramener à un niveau intermédiaire

Considérer les *k*-plus proches individus du nouvel individu à prévoir

Les *k* radios les plus proches, les *k* courriels les plus proches, etc.

& Fournir une probabilité

Difficultés

- Choix de la métrique (suppose représentation préalable)
- Choix de *k*

On compare un nouvel individu à k individus existants, difficiles à calculer :

Boîte noire prédictive très dépendante du jeu de données

Contraste fort avec les modèles statistiques : explicables, auditables
mais valables pour données « faciles » et moins puissants même dans ce cas

Importance des étapes de **préparation**

- Nettoyage du jeu de données
- Représentation des données de texte, d'images, vidéo, son
→ p.ex. par coefficients de Fourier

Piloter : couper le jeu de données en ensemble d'entraînement
+ ensemble de validation

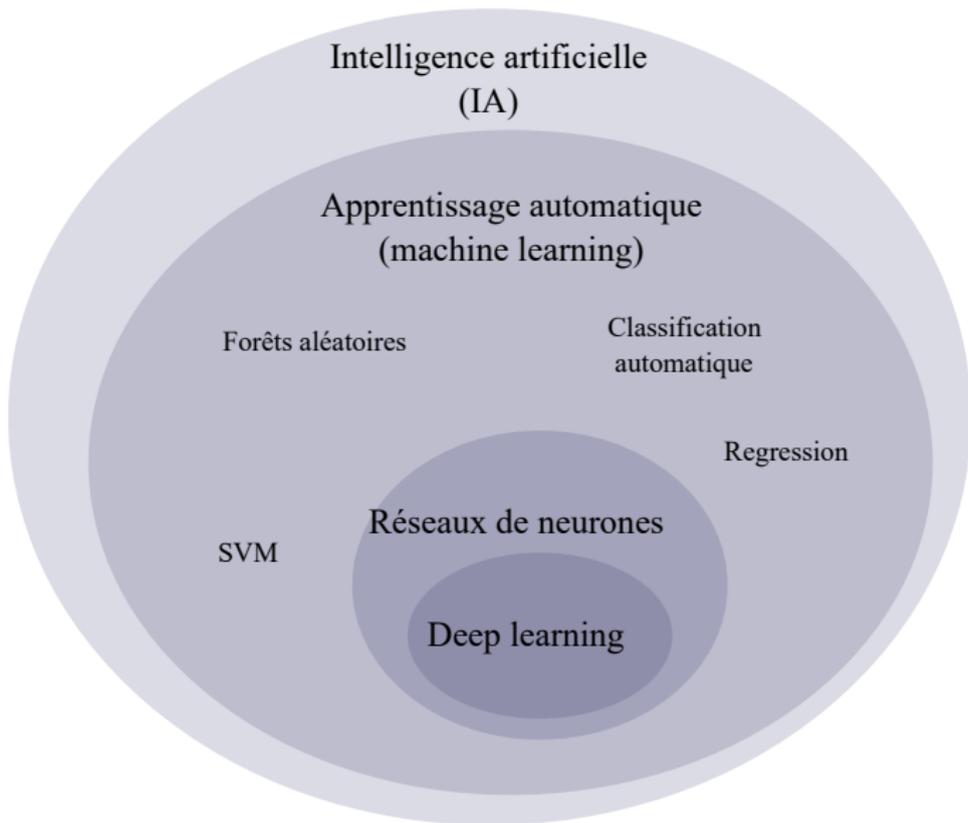
Un transparent sur le **vocabulaire** et les **buzz words**

Autres méthodes

- Arbres de décisions, forêts aléatoires
- Réseaux de neurones, **deep learning**

Différents ensembles

- Apprentissage automatique (**machine learning**) = les méthodes
- Data science = toute la chaîne de travail
- Big data = quand grand volume de données hétérogènes (vidéos, images, textes)
- Intelligence artificielle = tout ça, et plus encore



Source : Wikipédia

Progrès possibles

Aides au diagnostic médical, suppression des tâches répétitives,
aide mais pas substitution à la décision

Dangers redoutés

Monopolisation des données et captation de la plus-value en résultant, ex : GAFA
Dangers pour les libertés et l'autonomie des individus

Pour vos entreprises

Application nécessite **besoin identifié** et **maturité certaine**

Le cas échéant : stagiaires IUT, écoles d'ingénieur pour démarrer
Cf. toutes les méthodes mathématiques sont codées dans des « packages » libres

Au fait, mes travaux de recherche...

Agrégation de méthodes de prévision
(pour grands groupes ou entreprises spécialisées en prévisions)

Contrats de collaboration recherche en cours avec EDF et Cdiscount

Pour en savoir plus :

[http://www.hec.fr/Knowledge/Point-de-Vue/
Agreger-les-previsions-pour-preparer-l-avenir](http://www.hec.fr/Knowledge/Point-de-Vue/Agreger-les-previsions-pour-preparer-l-avenir)