HAL
open science

Les IntelligenceS ArtificielleS pour l'Industrie : quel type pour quelle innovation ?

Fabien Moutarde

To cite this version:

Fabien Moutarde. Les IntelligenceS ArtificielleS pour l'Industrie : quel type pour quelle innovation ?. Entretiens de Toulouse :la formation par le débat, Apr 2019, Toulouse, France. hal-02097636

HAL Id: hal-02097636

https://minesparis-psl.hal.science/hal-02097636

Submitted on 20 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Entretiens de Toulouse, 10-11 avril 2019

Les IntelligenceS ArtificielleS pour l'Industrie : quel type pour quelle innovation ?

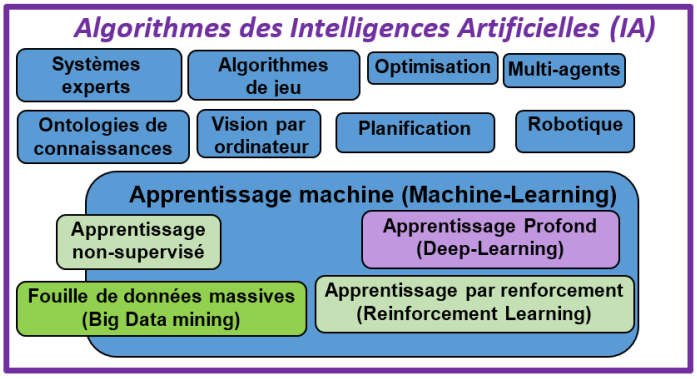
Fabien Moutarde, Centre de Robotique, MINES ParisTech, PSL Université
Fabien.Moutarde@mines-paristech.fr

Selon Wikipedia, l'Intelligence Artificielle (IA) est *« l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence »* (sous-entendu « humaine »). Mais l'intelligence elle-même est un concept assez mal défini et très polysémique. L'assimiler à la seule *capacité de raisonnement*, comme il vient en premier à l'esprit, serait très réducteur : nous ne sommes pas de purs esprits, mais des êtres situés dans un environnement où nous agissons, avec lequel nous interagissons, et aux changements duquel nous devons nous adapter en permanence. L'Intelligence (humaine) inclut donc en fait aussi :

1/ la faculté *d'analyser/interpréter une situation*, ce qui passe par une *Intelligence* que l'on pourrait appeler « *de perception* » (comprendre ce qui est autour) ;
2/ l'aptitude à anticiper par prévision des mouvements/actions autour (« *intelligence de prévision* »);
3/ la capacité à agir de façon pertinente et réactive, ce qui nécessite une sorte d'*intelligence « de comportement »* ;
4/ l'aptitude à communiquer ou au moins *interagir* ;
5/ et enfin, la faculté d'évoluer et de s'améliorer, ce qui nécessite un forme de *d'adaptabilité* et de *curiosité*.

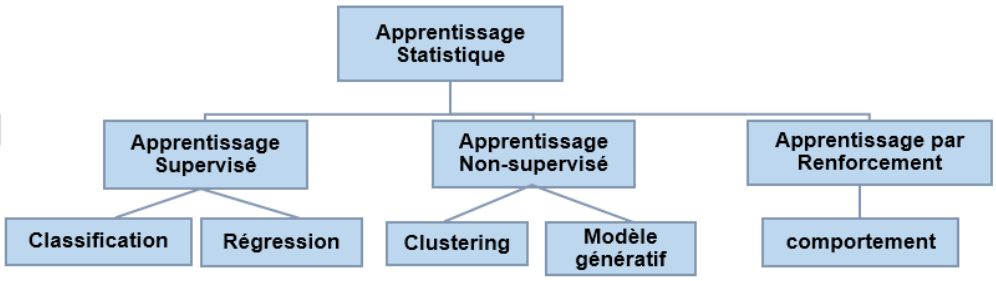
Les divers types d'Intelligences Artificielles

Ces divers volets de l'intelligence se retrouvent donc dans l'Intelligence Artificielle, comme on peut voir par exemple dans la « cartographie de l'IA » que l'INRIA a proposée dans son Livre blanc de l'IA [https://www.inria.fr/content/download/103897/1529370/AI_livre-blanc_n01.pdf]. La diversité et l'hétérogénéité de ce que recouvre l'IA est telle qu'il serait plus exact de parler DES IntelligenceS ArtificielleS. Car si les medias grand public ont une fâcheuse tendance par exemple à utiliser les termes IA, Deep-Learning, et Big Data de façon interchangeables et quasi-synonymes, l'IA est un « domaine valise » incluant des théories et techniques extrêmement variées, maîtrisées par des communautés scientifiques parfois même disjointes, et dont Deep-Learning et Big Data ne sont que des sous-parties.



Tous les sous-domaines de l'IA peuvent et vont contribuer aux transformations et aux innovations liées à l'essor actuel de la numérisation et de l'automatisation, qui touche tous les secteurs d'activités, y compris de plus en plus d'applications « grand public ». Mais en réalité, l'emballement actuel sur l'IA est principalement lié à des progrès récents et très significatifs dans le domaine de l'apprentissage machine (Machine-Learning), et en particulier de l'apprentissage profond (Deep-Learning).

Ce sont donc surtout les percées récentes en apprentissage statistique qui offrent de nouvelles opportunités d'innovations. Ce domaine est lui-même subdivisé en plusieurs sous-catégories, dont le point commun est de combiner modélisation statistique empirique avec les techniques d'optimisation numérique. On distingue en effet trois grands « types » de Machine-Learning : l'apprentissage *supervisé*, l'apprentissage *non-supervisé*, et l'apprentissage *par renforcement*.



L'apprentissage supervisé (le plus fréquemment utilisé), consiste à exploiter un nombre important d'exemples associés à des valeurs-cibles (target values) pour engendrer de façon semi-automatique un modèle *empirique* (par opposition aux modèles physiques, par exemple) de la forme y=h(x) effectuant une sorte d'interpolation-approximation capable de prédire par généralisation la valeur-cible associée à des inputs autres que ceux inclus dans la base. Quand les targets sont à valeurs continues, il s'agit de faire une *régression* ; d'ailleurs la régression linéaire par moindres carrés est en fait la forme la plus simple et la plus ancienne d'apprentissage statistique supervisé. Si les valeurs sont discrètes, elles correspondent généralement à des labels de « catégories », et le résultat de l'apprentissage supervisé sera un modèle permettant de *classifier* automatiquement de futurs inputs.

L'apprentissage non-supervisé consiste à analyser des données *sans valeurs-cibles associées*, pour modéliser leur répartition dans l'espace des inputs possibles : c'est essentiellement de la fouille de données (datamining), comme par exemple le « clustering » qui consiste à partitionner les exemples en plusieurs ensembles ayant des points communs qui les distinguent de ceux des autres « clusters », ou encore un modèle génératif ayant la même distribution de probabilités que les données.

Enfin, *l'apprentissage par renforcement* (Reinforcement Learning, RL) consiste typiquement, par des modifications itératives en fonction de « récompenses » (reward) selon le résultat, à produire un comportement automatisé optimum afin de réaliser une certaine tâche sans avoir besoin de la programmer explicitement. Par exemple, si un robot manipulateur doit attraper toutes les pièces sur une table située devant lui, et les placer dans un réceptacle, il suffira d'abord de définir une pénalisation (récompense négative) pour chaque pièce « oubliée », et une récompense importante pour chaque pièce déposée au bon endroit. Ensuite l'algorithme d'apprentissage par renforcement effectuera des milliers (voire des millions !) de tentatives (le plus souvent en simulation, pour que cela puisse être totalement automatique) entrecoupées d'améliorations du modèle d'actions, calculées pour augmenter le cumul de récompenses, jusqu'à obtenir un comportement adéquat et robuste.

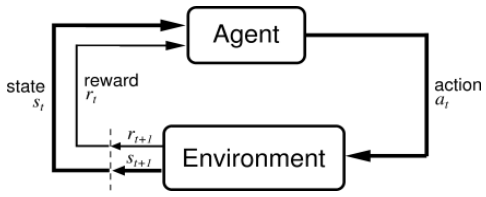


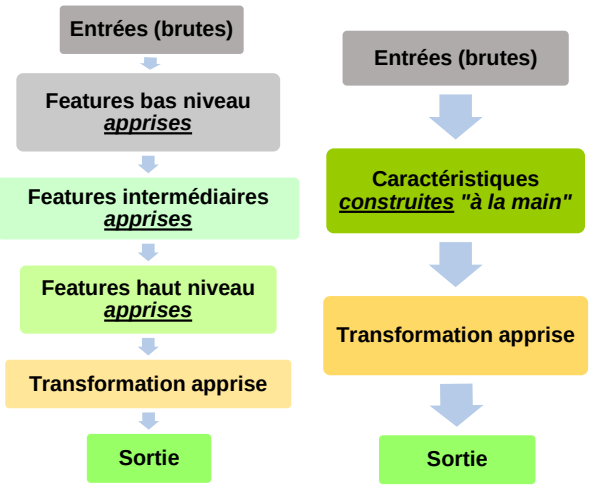
Schéma de principe de l'apprentissage par renforcement. L'agent optimise sa « politique » $a_t=\pi(s_t)$ pour maximiser sa récompense cumulée.

Exemple d'application du RL pour apprendre un comportement à des robots.[1]

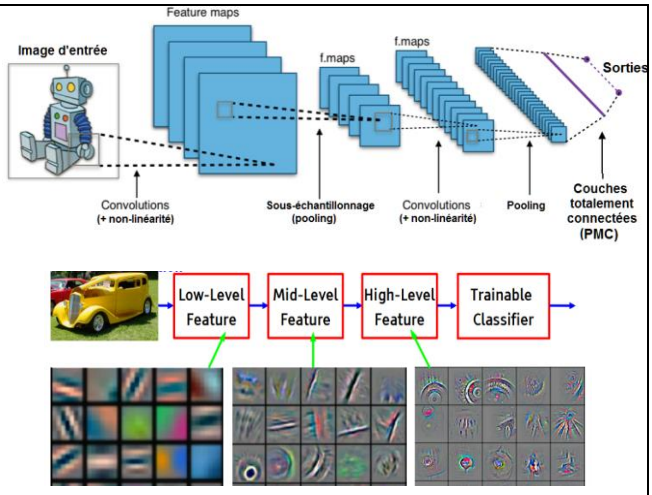
Quant à l'apprentissage profond (deep-learning), cela correspond à un type particulier de modèle, indépendamment de la nature supervisée ou non de la tâche d'apprentissage. Il se caractérise par des architectures consistant en une série de transformations successives à partir des données brutes, jusqu'à une représentation simplifiée qui rend la tâche finale (par exemple classification) facile à effectuer, et qui *a posteriori* s'avère assez « sémantisée ». Cette approche a montré sa puissance en particulier pour l'analyse des images, avec les réseaux convolutifs (Convolutional Neural Networks = ConvNets ou CNN) qui transforment par filtrages et sous-échantillonnages successifs l'image initiale en une hiérarchie de représentations de plus en plus « abstraites ». Ce dernier domaine est un de ceux où les progrès récents ont été les plus spectaculaires. Des variantes non-supervisée (réseaux adversariaux génératifs = Generative Adversarial Networks, GAN ; et/ou auto-encodeurs variationnels) de ces réseaux convolutifs ont aussi

[1] Image issue de http://robohub.org/deep-learning-in-robotics/

récemment montré une bonne capacité à apprendre des « modèles génératifs » capables de produire des images fictives très réalistes après apprentissage de la probabilité de distribution.



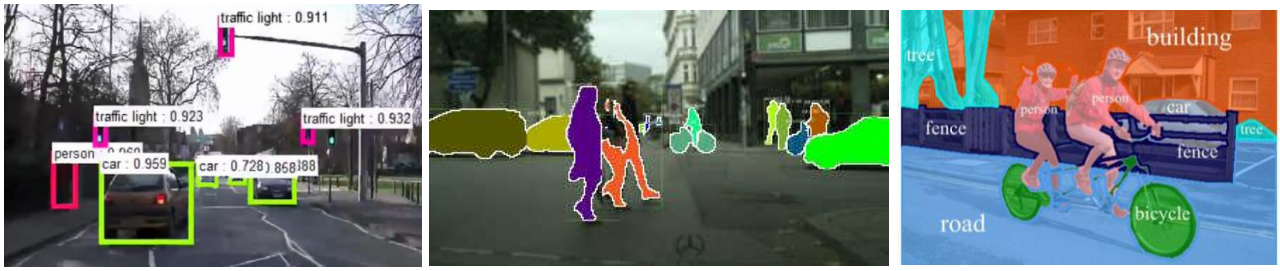
Chaîne de traitement de l'apprentissage profond (à gauche) comparée à celle usuelle avec les techniques d'apprentissage « classiques » (à droite).



En haut, architecture typique d'un réseau neuronal convolutif.[2] *En bas, aspects typiques des filtres appris pour les couches successives d'un réseau convolutif.*

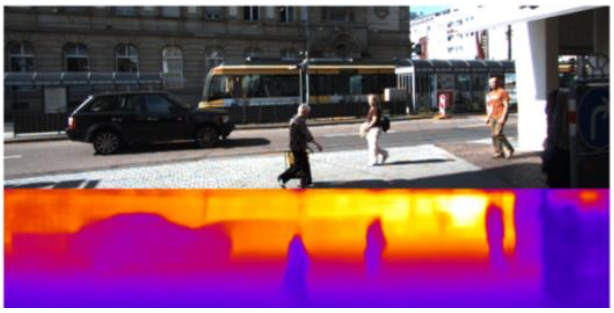
Le même principe de convolutions successives a été aussi appliqué non plus sur des images, mais dans le domaine du traitement automatique de la langue naturelle (TALN, Natural Language Processing, NLP) : les inputs sont alors des phrases, vues comme des séquences de mots, et les convolutions traitent de petits groupes de mots consécutifs. Ceci, associé à des architectures de réseaux récurrents, a permis une importante percée en traduction automatisée et compréhension de texte. Enfin, le monde entier a été frappé par la victoire en 2017 du programme AlphaGO contre le champion du monde (humain !) du jeu de Go pourtant d'une grande complexité stratégique. Ce résultat a illustré de façon éclatante le potentiel des algorithmes d'apprentissage par renforcement pour engendrer des comportements sophistiqués.

Comme évoqué précédemment, les potentiels de l'IA pour des innovations, en général comme dans l'Industrie, sont principalement liés à l'apprentissage statistique, et notamment aux avancées les plus récentes et significatives du domaine. Ces dernières sont manifestement d'abord dans le domaine de *l'intelligence de perception, en particulier l'analyse « sémantique » de scène visuelle par des réseaux convolutifs profonds*. Ceux-ci permettent en effet maintenant de produire facilement et en temps-réel sur des vidéos une annotation automatisée des emplacements et types d'objets dans la scène, voire même une segmentation sémantique partielle ou complète (comme illustré ci-dessous). Mieux encore, de nouvelles fonctionnalités apparaissent désormais chaque année, comme par exemple la détermination automatique des postures des humains (squelettisation) ou encore l'estimation des distances de chaque pixel (calcul d'une image de profondeur approximative).



« Sémantisation » d'images par réseau convolutif : à gauche, boites englobantes des objets de plusieurs catégories ; au centre, localisation par masques de détourage ; à droite, segmentation sémantique de la totalité de l'image.

[2] Schéma adapté de http://fr.m.wikipedia.org/wiki/Fichier:Typical_cnn.png



Percées récentes des réseaux convolutifs profonds : à gauche, inférence des postures humaines ; à droite, estimation approchée des distances (image de profondeur en dessous).

Ces nouvelles techniques d'interprétation sémantique d'image en temps-réel ouvrent d'innombrables possibilités, tant en termes applicatifs (juste utiliser les réseaux convolutifs « sur étagère » pour fournir des informations sémantiques en entrée d'un autre algorithme) qu'en terme de R&D : tout comme les postures et la 3D, il est probablement possible d'apprendre à inférer de nombreuses autres choses utiles par un réseau convolutif appliqué à des images. Comme on le voit sur les exemples de résultats, l'exploitation applicative s'est jusqu'ici beaucoup focalisée sur les scènes extérieures (avec en « ligne de mire » les Véhicules Autonomes), ou sur des intérieurs type domicile/bureaux (pour la robotique de service). Mais ***des résultats similaires (quitte à adapter un peu les classes et « fine-tuner » un réseau convolutif pré-appris sur des bases d'images générales) sont très certainement atteignables dans des environnements industriels.***

La deuxième catégorie d'IA où des progrès très importants sont apparus est celui de *l'intelligence d'interaction, en particulier la reconnaissance vocale, et l'analyse/traduction/compréhension du langage naturel*. Les « enceintes intelligentes » Amazon Echo™, Google Home™ et HomePod™ d'Apple qui se disputent le marché des particuliers, constituent un parfait exemple de nouveau produit ainsi rendu possible. Ces mêmes fonctionnalités peuvent certainement s'avérer extrêmement intéressantes pour développer des interfaces « naturelles » pour des machines et robots. D'autant plus que l'application de l'apprentissage profond sur les trajectoires dans l'espace des articulations (poignets, coudes, etc…), maintenant détectables dans des images ordinaires (cf. plus haut) va certainement permettre des avancées spectaculaires dans le domaine de la reconnaissance de gestes et d'actions.

Enfin, des pistes plus amont, mais à très forts potentiels sont ouvertes :

- par l'apprentissage non-supervisé de *modèles génératifs*, par exemple pour améliorer la *prévision automatisée des mouvements/actions* autour de soi
- par *l'apprentissage par renforcement* (probablement combiné avec l'apprentissage « par démonstration ») pour produire automatiquement des *comportements adaptatifs et optimisés*, par exemple pour des robots collaboratifs, mais pourquoi pas aussi globalement sur l'ensemble d'une chaîne de production

Une tentative de synthèse de ces potentialités est proposée dans le tableau suivant :

Type d'IA	**Fonction visée**	**Technique à utiliser**
Analyser	Analyse visuelle de scène	*Réseau convolutif profond (ConvNet)*
Anticiper/prévoir	Prévision de mouvement et action	*Régression* *Modèle génératif ?*
Ré(agir)	Aide à la décision	*Classifieur (quelconque)*
	Comportement adaptatif	*Reinforcement Learning (RL)*
Interagir	Interface Machine « naturelle » (voix et gestes)	*Speech-to-Text* *+ Deep-Learning(NLP)* *ConvNet(mouvement)*

On voit donc que le type d'algorithme d'IA potentiellement intéressant peut être très différent selon la fonction recherchée. Intuitivement, c'est probablement en combinant plusieurs des fonctions ci-dessus que les innovations les plus « disruptives » (et donc les plus intéressantes !) sont susceptibles d'émerger. Par exemple, en intégrant dans un seul robot collaboratif à la fois les algorithmes les plus avancées (réseaux convolutifs profonds) d'analyse sémantique de scène visuelle, des capacités de prévisions des mouvements et actions, ainsi qu'un comportement adaptatif obtenu via l'apprentissage par renforcement, et l'aptitude aux interactions « naturelles » par voix et gestes désormais atteignable par apprentissage profond. Un tel robot collaboratif peut constituer pour des opérateurs humains un véritable compagnon de travail polyvalent et

adaptatif, une sorte de « super-outil » augmentant la productivité sur les chaînes de montage tout en diminuant la pénibilité des postes. Ou peut aussi être un produit destiné à l'industrie des services (commerces, restauration, etc…).

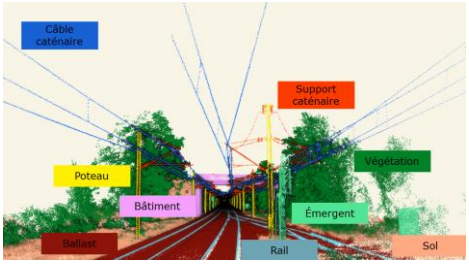
Un exemple de robot collaboratif (Baxter) déployé dans une usine.[3]

Robot logistique d'Exotec (PME française), déjà déployé dans un entrepôt de Cdiscount.

Exemple d'intégration de nombreuses briques de divers types d'IA dans un robot interactif polyvalent

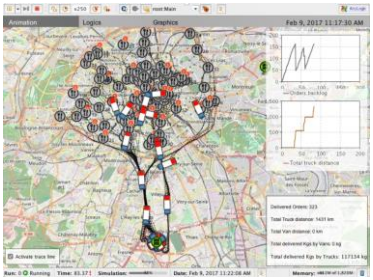
Exemple de commande gestuelle d'un robot (ici avec un capteur musculaire de type EMG).[4]

Plus généralement, quelles innovations sont rendues possibles dans l'Industrie par les divers types d'IA ? Tous les métiers sont potentiellement concernés : non seulement le *processus de fabrication (robots collaboratifs, etc…)*, mais aussi en amont *l'analyse de la demande du marché* et la conception de *nouveaux produits*, et en aval la *vente (chatbots, segmentation de clientèle,…)*, la *logistique*, et le service après-vente (*maintenance prédictive*). A chacun, à partir de ses propres « connaissances métier », et en ayant ou acquérant une compréhension minimum des potentialités des divers types d'IA, d'imaginer pour son secteur et métier soit des améliorations significatives de l'existant, soit des produits ou des process radicalement différents et disruptifs, rendus possibles par les avancées récentes en IA/Machine-Learning.



Cartographie digitale : numérisation 3D + sémantisation du nuage de points

Concept de voiture volante d'Airbus et Italdesign



IA pour la logistique urbaine

[3] Image issue du site http://www.futurism.com

[4] Image issue de https://spectrum.ieee.org/automaton/robotics/robotics-hardware/video-friday-rescue-quadruped-gesture-controlled-robot-arm-self-driving-van-1986