Association Nationale des Directeurs des Systèmes d'Information
www.andsi.fr

Intelligence Artificielle – Point d'étape et pratique

Compte rendu de la présentation du 11 septembre 2018, à La Terrasse

Compte rendu rédigé par Laure MUSELLI & ANDSI

En bref…

Pierre DELORT, DSI et Professeur Invité à Télécom ParisTech, propose un état des lieux des pratiques et de la recherche autour du Big Data, du machine learning et de l'intelligence artificielle, des notions aujourd'hui largement mobilisées, porteuses d'espoirs, voire de fantasmes pour les entreprises, et pourtant encore assez mal définies et différenciées. A travers des exemples concrets et une démonstration de création, entrainement et test d'un réseau de neurones profond, il présente les outils permettant de rechercher des modèles et revient également sur les enjeux pratiques de prises de décisions sur la base d'algorithmes de moins en moins compréhensibles par l'homme.

Big Data, Machine Learning et Intelligence Artificielle ?

« If your torture the data enough, nature will always confess » (Coase R., 1981)

« Vous avez le pouvoir magique d'évoquer de votre baguette des palais de féerie. Moi, je suis le modeste ouvrier des cathédrales, qui apporte obscurément sa pierre à l'édifice auguste qu'il ne verra jamais. Au premier souffle de la réalité, le palais de féerie s'envole, tandis qu'un jour la cathédrale républicaine lancera sa flèche dans les cieux. ». (Clemenceau G., 1906).

« The current debate on "the future of work" or "jobs at risk of automation" seems to implicitly adopt a pure science-push view, which assumes a path for technology driven by what science makes achievable, rather than what is needed by firms » (Presidente G., OCDE, 2017)

La première citation donne le cadre, les données, la seconde illustre le fantasme existant autour du Big Data ou de l'intelligence artificielle, alors que la troisième rappelle que ce sont les firmes qui décideront de d'utilisation de ces technologies.

A quoi fait-on référence lorsque l'on parle de Big Data, de machine learning ou d'intelligence artificielle ?

Big data et recherche inductive :

Le **Big Data** consiste à **créer, en exploratoire et par induction, sur des masses de données à faible densité en information, des modèles à capacité prédictive, sous la condition que le futur soit semblable au passé**.

Le cas de la prédiction de l'évolution de l'épidémie de grippe à partir des recherches Google constitue une bonne illustration de ce principe.

Aux Etats-Unis, afin de mettre en place les outils de maîtrise sanitaire en fonction de l'évolution de l'épidémie de grippe, le Center for Disease Control (CDC) dispose d'un réseau appelé Sentinel de médecins déclarent les cas de grippe constatés chez leurs patients. Ce dispositif fonctionne correctement mais souffre d'un retard de 10 jours environ entre les premiers symptômes et la déclaration par le médecin, du fait des délais de prise de rendez-vous chez le médecin ou de déclaration par ce dernier. De ce fait, l'objectif du CDC consiste à mettre en place des dispositifs permettant de réduire ce délai et de suivre le plus précisément possible l'évolution de l'épidémie.

Des chercheurs ont mis en évidence la capacité à prédire l'évolution de l'épidémie à partir des recherches Google effectuées. Parmi les 50 millions de termes de recherche, ils ont identifié les 45 mots-clés ayant le plus fort coefficient de corrélation (coefficient de Pearson) avec les données historiques du CDC, parmi lesquels « cold », « flu », « how to treat flu ». Sur ces 45 mots-clés, ils ont ensuite réalisé une régression logit.

On constate que les prévisions Google sont alignées avec les données réelles de l'épidémie de grippe, et permettent de gagner 10 jours environ sur les informations de Sentinel. Toutefois, la capacité prédictive est conditionnée à une **hypothèse de futur semblable au passé**, qui n'est pas toujours vérifiée, comme le prouve, entre autres, une alerte sanitaire à New-York en 2013.

Le Big Data repose donc sur un **raisonnement de type inductif**, par opposition à un raisonnement déductif. Des requêtes sont tapées, des pages servies, ces informations (adresses IP, différentiation des « long clicks » et « short clicks » pour évaluer de la pertinence du résultat) sont conservées et anonymisées. Des technologies permettent ensuite d'inférer, sur ces masses de données, des modèles à capacité de forecasting ou *nowcasting*, en prise d'avance sur le futur ou le présent.

Relativement à la fouille de données (datamining), ce qui caractérise le **Big Data**, est que le **volume adressable important** (peut-être de l'ordre de l'exaoctet) permet d'exploiter de nouvelles classes de donnée, des **données à faible densité en information**, ou **clairsemées (cf. matrices creuses), pour y trouver des modèles absents à la conception**, les signaux faibles d'un monde dont la densité numérique devient significative. Ce sont les capacités actuelles en termes de processeur et de stockage qui permettent, grâce au traitement de gros volumes de données, de trouver des modèles imprévus dans une représentation du monde. Par exemple, les logs d'un ascenseur peuvent permettre de révéler des modèles de fonctionnement dans un immeuble, alors que ceci n'avait pas été prévu à l'origine.

Par comparaison, dans le cas d'un système d'information, on trouve plutôt un faible volume et une forte densité en information. Dans le cas des bases de données de courtiers de données tels qu'Acxiom, qui possède des bases de plus de 500 millions de personnes, on est sur un fort volume et une forte densité.

Finalement, **alors que l'informatique consiste à développer des modèles de données pour réduire le monde, dans le cas du Big Data, l'approche est différente, car il s'agit de chercher un modèle dans les données**.

Les outils de recherche de modèles

Il existe plusieurs types d'outils de recherche de modèles.

Les régressions de l'économétrie, qu'elles soient de type linéaire, logarithmique-linéaire, polynomial ou logit, elles permettent de prolonger une tendance et de cadrer un peu le futur. Peut-on parler d'intelligence artificielle pour y=ax+b ? Beaucoup, dont le conférencier, en douteraient.
Par exemple, on peut effectuer une régression sur les résultats scolaires, relativement aux revenus des zones scolaires. On peut en effet penser que plus les gens sont riches dans une zone, plus leurs enfants auront de bons scores à des tests. Il s'agit d'une forme d'induction.

Le plus grand danger de ce type d'outil est le sur-échantillonnage, qui permet de faire dire n'importe quoi aux données. Un article de 1995 avait en effet essayé de voir s'il était possible de prévoir l'indice Standards&Poor (S&P500) à partir de la production de beurre au Bangladesh. Il y arrivait avec un certain succès (R^2 de 0,78). Si on rajoutait à la production de beurre au Bangladesh, celle des États-Unis, ainsi que la production de fromage aux États-Unis, on améliorait la précision du modèle. On l'améliorait encore en rajoutant le nombre de moutons. Maintenant, pourquoi aller se fatiguer à compter des moutons au Bangladesh, alors qu'il est possible de donner le S&P500 à partir de l'année… dans l'échantillon. Hors de l'échantillon, le modèle ne fonctionne plus du tout.

Le **machine learning** renverse le principe traditionnel selon lequel on applique un programme à des données pour obtenir un résultat, puisqu'il **consiste, à partir des données et de leur output, à sortir un programme, c'est-à-dire un algorithme**. Il s'agit toujours d'induction.
Parmi les outils du machine learning, on trouve le SVM (Support Vector Machine), le Random Forest, le K-means….

Ces outils peuvent être classés en trois catégories :

- **supervisé**, le plus utilisé, qui consiste à étiqueter les données, par exemple en indiquant, pour un grand nombre de photos, s'il s'agit d'une photo de chat ou pas, pour que le programme puisse ensuite apprendre à reconnaître, parmi les photos, celles qui sont des photos de chats et celles qui ne le sont pas.
- **non-supervisé**, où le programme identifie par exemple une taxonomie (agrégats de points…). Il peut d'ailleurs être combiné à du supervisé, comme dans le cas d'identification de fraudeurs. En non-supervisé, on identifie des entreprises aux caractéristiques communes, puis on passe à du supervisé dans un classifieur binaire pour obtenir leur probabilité d'être fraudeuses ou pas.
- **renforcé**, moins utilisé, dont le principe consiste à fixer des règles. Par exemple, dans le cas de la voiture autonome, la règle peut être de ne heurter aucun objet.

Aucune définition précise de l'**intelligence artificielle** n'a émergé de la Commission Villani, qui a auditionné P. DELORT. On peut toutefois penser que les réseaux de neurones y correspondent bien. Par exemple, avec beaucoup d'images et de bons algorithmes, il était possible à une machine de reconnaître des objets, tels un type de chaussure (démonstration de Y LE CUN - passée en séance). Le réseau de neurones profond utilisé était probablement constitué de dizaines de couches cachées, dont une permettant de repérer les arêtes, l'autre les séparations, etc…

Démonstration autour d'un réseau de neurones profonds

Un **réseau de neurones est dit profond lorsqu'il est composé de plusieurs couches cachées**. On parle également de **deep learning**. Le but consiste à ce qu'un **algorithme crée tout seul des variables indépendantes qui n'ont pas été fournies**.

Dans la démonstration réalisée en séance, un réseau de neurones est créé dans le but, à partir de données, de détecter les cas de fraude.
Les données utilisées ont été téléchargées sur Kaggle, qui est une plateforme organisant des compétitions de sciences de données. Il s'agit de 280 000 transactions (anonymes) de carte crédit, chacune décrite par 29 descripteurs, plus le montant. La classe notée 0 indique qu'il s'agit d'une transaction normale et la classe notée 1 indique une fraude. 200 000 des 280 000 transactions vont être utilisées pour entraîner le modèle, qui sera testé sur les 80 000 transactions restantes. L'algorithme sera in fine capable donner aux transactions qu'il n'a jamais vues un degré de suspicion.

Ce réseau de neurones profond a été créé avec des modules Python et TensorFlow, un outil Google open source.

Le data set est entré, puis les réglages effectués. Il s'agit de définir les caractéristiques d'entraînement que le réseau de neurones va subir, ici entre autres :

- 20 passes d'entraînement par lots de 128 transactions ;
- Le choix du mode de descente de gradient (descente stochastique) ;
- Le poids des classes a été ici changé en raison du grand déséquilibre des données….

Le réseau de neurones créé possède finalement 870 paramètres pour la couche d'entrée, 3480 pour la deuxième couche, 16 000 pour la troisième couche, et pour la couche finale, 129 paramètres, soit en tout 21 000 paramètres, auxquels il va falloir trouver, pour chacun, une valeur. C'est à cet effet que le réseau de neurones est entrainé au cours des 20 passes prévues, pour créer ces paramètres et donc in-fine le modèle. On obtient ainsi un classifieur binaire de fraude ou non-fraude caractérisé par une perte (c'est-à-dire la distance entre les points et le modèle) de 0,009 et une précision de 99,94%.

De l'explicabilité des algorithmes

Dans le cas d'une classification ordinaire sur un sujet non-critique de type « fraude vs. pas fraude », on peut se satisfaire d'indicateurs globalement assez bons sans que cela ne soit problématique. En revanche, sur des sujets critiques avec une forte asymétrie (il est relativement facile de passer de la vie à la mort, l'inverse est plus compliqué), les indicateurs statistiques deviennent cruciaux. La question de **l'insertion de ce type d'algorithme complexe dans les mécanismes de décision des entreprises** relève ainsi d'une vraie problématique aujourd'hui, car ceux-ci doivent **rester compréhensibles des décideurs**. C'est ce que l'on appelle l'**explicabilité des algorithmes**.

Il existe aujourd'hui une vraie controverse autour du danger à laisser le jugement aux algorithmes, même s'il a été mis en évidence dans les processus de jugement humain un effet halo donnant lieu à des raisonnements biaisés et des jugements conditionnés par la première appréciation. Y. LE CUN, avance l'argument que des médicaments sont

validés alors que leur fonctionnement n'est pas compris. Pourquoi ne pas utiliser les mêmes process de validation pour des algorithmes ?

Mais le sujet-là reste : « **peut-on mettre en place des algorithmes dont on ne comprend pas le fonctionnement ?** ». Il n'est pas évident de comprendre le modèle développé par un réseau de neurones, même très petit, avec ses 21 000 paramètres. Se pose aujourd'hui une **question de responsabilité**, qui est le véritable enjeu de l'insertion des algorithmes dans les entreprises. Une personne doit être responsable de la décision.

A ce sujet, un article de recherche relate le test par un hôpital de trois solutions destinées à prendre les décisions liées au placement en soins intensifs ou pas de patients souffrant de pneumonie : un système à base de règles, une régression logistique et un réseau de neurones. Le réseau de neurones, qui avait été entraîné sur une douzaine de milliers de cas et une quarantaine de descripteurs de l'état du patient, renvoyait systématiquement chez eux les asthmatiques, alors que leur mortalité est bien supérieure au 11% sans. Cela s'explique par le fait que le réseau de neurones avait été entraîné avec des données indiquant que les asthmatiques avaient une faible mortalité en cas de pneumonie car ils avaient justement été conduits systématiquement et directement en soins intensifs. Dès lors il avait décidé d'utiliser la régression logistique certes moins performante que le réseau de neurones (AUC 0,77 vs 0,86) mais plus compréhensible, les médecins se disant qu'ils avaient pu éviter de « tuer » des asthmatiques et qu'ils ne souhaitaient pas que cela se reproduise sur d'autres populations à risque par manque d'intelligibilité de l'algorithme.

Perspectives partielles pour une meilleure explicabilité… et dangers…

Aujourd'hui, des recherches visent à améliorer l'explicabilité des algorithmes. Cela peut passer par l'ajout d'un second réseau de neurones. Par exemple, dans le cadre d'une étude portant sur la reconnaissance d'espèces d'oiseaux, un premier réseau de neurone fonctionne en mode supervisé pour classer les photos. Un second réseau de neurones est, lui, réglé de façon à générer des phrases « expliquant » le classement de la photo, la fonction de perte récompensant la discrimination entre classes. De ce fait, l'inintelligibilité des algorithmes peut être en partie palliée.

Beaucoup de chercheurs ou praticiens en intelligence artificielle, placent le **curseur de l'intelligence artificielle au niveau d'un algorithme non-intelligible et non-modulaire**, en analogie avec le cerveau humain. Ce qui s'applique bien à un **réseau de neurones incluant plusieurs millions de paramètres.**

Parmi les recherches en cours, on trouve également l'« adversarial neural network », qui travaille sur des réseaux de neurones réglés en fonction de différences non visibles par un humain, afin d'influencer les résultats produits par les réseaux de neurones de référence. Par exemple, des réseaux de neurones sont spécialement réglés pour tromper le réseau de neurones en charge de la reconnaissance d'un panneau de signalisation. Ce type de recherche fait partie des dangers pointés par les spécialistes de l'intelligence artificielle.

L'avenir dans les nuages

Finalement, on peut penser que le cloud prendra un rôle croissant dans l'Intelligence artificielle compte tenu de sa formidable capacité à rassembler des données et à les faire étiqueter gratuitement par les utilisateurs.

En ce qui concerne les outils de machine learning, Google a développé TensorFlow, disponible sous licence open source. De leur côté, AWS et Azure proposent Apache MXNet, également open source.

Google Cloud Platform permet de réaliser des opérations de reconnaissance de photo (picture to text), par exemple à partir d'une photo trouvée sur le net, enrichit avec Wikipédia le texte (Pour « Villepinte » reconnu sur une affiche, propose l'article Wikipedia.

De la même façon, si Google Cloud Platform, ne permet pas de reconnaître le modèle ni même le type de chaussure à partir d'une photo réalisée avec un smartphone (en tout cas avec celui du conférencier), il permet de reconnaître, à partir d'une photo d'une plante (du jardin d'un de ses amis), qu'il s'agit d'une plante (probabilité de 90%), voire même le type de plante (Alismatales) et ce avec une probabilité de 70%.

Débat

Intervenant : Finalement, la reconnaissance d'image, c'est ni plus ni moins que ce que fait notre cerveau depuis toujours. Trouver que c'est qu'une plante, n'importe quel botaniste sait le faire.

Pierre Delort : C'est beaucoup moins cher, beaucoup plus rapide. C'est un peu comme Shazam : avant Shazam, pour reconnaître une chanson il fallait chanter ce qu'on avait entendu à un disquaire, ce qui ne fonctionnait pas (en tout cas pour moi), ou alors se débrouiller pour obtenir la programmation de la station de radio…

Int. : Je ne suis pas en train de minimiser l'importance de l'analyse d'une grande quantité de données, mais toutes ces technologies de recherche dont tu parles, ce sont des technologies des années 50 – 60.

P. D. : Les réseaux de neurones, oui. C'est vrai. La différence, c'est l'ouverture à des millions de photos étiquetées. Et la technologie qui permet de les traiter.

Int. : On parle beaucoup de traitement d'images, mais est-ce vraiment utilisable dans un contexte d'entreprise ?

Int. : Nous avons tous des systèmes de LAD RAD (lecture automatique de documents) chez nous. Ça prend un temps fou et ça n'a rien à voir avec ces technologies-là. Les algorithmes commencent à poindre et ne sont pas encore satisfaisants. Pour reconnaître une feuille de soins, en santé, on a l'impression que c'est facile, mais ça reste encore terriblement compliqué. Je pense que ces technologies-là vont nous faire progresser de manière incroyable.

P. D. : On peut aussi citer le cas du concours Engie, qui avait pour but de cadrer la gestion d'une éolienne, afin de voir si elle se situe dans une zone correcte de fonctionnement, s'il faut en changer la conduite ou l'arrêter pour réparation ou maintenance. L'idée était d'avoir, à partir d'une soixantaine de variables indépendantes comme la température de l'air, la vitesse ou la température interne de l'éolienne, la production électrique théorique de l'éolienne, à comparer avec sa production réelle. C'est un modèle qui cadre le physique (digital twin) et qui dit : « le fonctionnement est conforme » ou « le fonctionnement n'est pas conforme ».

Int. : Peut-être à titre d'exemple, dans l'assurance, nous travaillons, en ce moment, avec des start-up sur un projet visant, à partir de photos de voitures accidentées, à identifier immédiatement si elle est réparable ou non, ce qui n'est pas si facile, mais surtout à évaluer le coût de la réparation pour indemniser quasiment en temps réel et orienter vers un réparateur. C'est assez compliqué car il y a beaucoup d'images, mais selon la façon de prendre la photo, le nombre de photos prises sur le véhicule, nous venons de faire des tests montrant que ce type d'identification est possible. Pour l'instant, ça fonctionne bien sur les accidents de l'avant droit. Donc ça reste limité et il y a quelques start-up qui travaillent là-dessus. Nous avons une application très concrète avec laquelle nous pensons travailler d'ici quelques mois, afin de dire immédiatement, à partir d'une photo de véhicule accidenté prise avec un téléphone, s'il est épave ou pas et s'il n'est pas épave, travailler sur l'évaluation des réparations et l'orientation vers des réparateurs. Ça, c'est très concret dans l'assurance.

P. D. : Il faut bien voir que le challenge consiste à insérer l'algorithme dans un process. Je pense que ce qui est important est la rapidité de mise en œuvre. Le process va étiqueter les résultats ou une partie des résultats pour savoir si l'algorithme, qui se base sur le passé, n'est pas trop situé dans le passé. Il faut toujours vérifier si l'algorithme se trouve toujours dans la bonne zone de fonctionnement et, si ce n'est pas le cas, soit le restructurer, soit le réentraîner, soit le changer. Cela peut constituer des années de maintenance à assurer, et je pense que les DSI doivent être assez attentifs à insérer les algorithmes dans des process, mais aussi à avoir des dispositifs pour maintenir l'algorithme. Ainsi, dans une entreprise de Telco, l'algorithme de détection de fraude à la souscription est réentraîné tous les mois environ.

Int. : De mon côté, j'ai un cas, mais qui en est vraiment au tout début. Nous fabriquons des planches de bord, c'est-à-dire la partie qui est derrière le volant. Sur les modèles de luxe, quand vous appuyez sur cette partie, c'est en général mou. C'est mou parce qu'il s'agit d'une peau derrière laquelle on injecte de la mousse selon un processus très instable qui s'appelle le « moussage », et sur lequel nous avons énormément de rebut. Ce processus est instable non seulement parce que la mousse ne se répartit pas correctement, mais aussi parce qu'il implique énormément d'intervention humaine, ce qui rend les choses assez compliquées. On se demandait si la solution pouvait relever du data mining, du Big Data ou éventuellement du machine learning. C'est simplement une façon de poser le problème et d'ailleurs, je vais essayer de mettre un stagiaire sur ce sujet.

P. D. : La définition la plus simple, je pense, mais aussi la plus éclairante, c'est que le machine learning, c'est ça. Apparemment, ton exemple est un classifieur binaire. Tu as un certain nombre de variables indépendantes qui vont être, par exemple, la température d'injection, la pression, la température externe, éventuellement caractériser l'opérateur, son niveau de formation ou son expérience, puis tu vas qualifier le résultat ; « c'est bon » ou « ce n'est pas bon ». Tu es bien en Machine Learning supervisé, parce que tu vas dire devant le résultat: « c'est bon », « ce

n'est pas bon ». Puis, le réseau va trouver l'algorithme correspondant qui te permettra de trouver les réglages corrects.

Int. : Mais tu confirmes que c'est effectivement plus du machine learning que du big data ou du data mining, a priori ?

P. D. : Chacun a un peu sa définition, mais je pense qu'ilest solide de dire que c'est du machine learning parce que ta machine va créer un algorithme. Donc, cela va te trouver ton algorithme entraîné sur tes données. Cela serait du Big Data, avec beaucoup de données, et principalement un sujet (mathématique) de matrice creuse, mais cela ne semble pas être le cas.

Int. : En ce qui concerne la quantité de données, justement. A partir de quelle quantité de données le machine learning est-il efficace ?

Int. : Le volume de données dépend aussi du nombre d'occurrences, du pourcentage de cas de fraudes par exemple. Celui-ci étant faible, tu as besoin de beaucoup de données, de beaucoup d'occurrences.

P. D. : D'une manière générale, plus tu as de données, mieux c'est. Maintenant, il ne faut pas qu'il y ait sur-échantillonnage et le modèle doit être généralisable sur toutes les données. En finance, par exemple, pour les algorithmes de high frequency trading, on privilégie une quantité importante de données et des algorithmes plutôt simples, mais maîtrisables et fréquemment entraînés.

Présentation de l'orateur

Pierre DELORT est DSI groupe et professeur invité à Telecom ParisTech. Il a connu une carrière de consulting (Gemini Consulting) en operation management, puis de DSI (RFF, Inserm…). Il est l'auteur, aux Presses Universitaires de France, de « Le Big Data », dans la collection *Que Sais-je ?* .