Intelligence artificielle, Machine Learning et Big Data : de quoi parle-t-on?

Autor(en): **Cudré-Mauroux, Philippe**

Objekttyp: **Article**

Zeitschrift: **Kriminologie / Schweizerische Arbeitsgruppe für Kriminologie SAK = Criminologie / Groupe Suisse de Criminologie GSC = Criminologia / Gruppo Svizzero di Criminologia GSC**

Band (Jahr): **38 (2021)**

PDF erstellt am: **13.09.2024**

Persistenter Link: https://doi.org/10.5169/seals-1051601

Intelligence artificielle, *Machine Learning* et *Big Data* : de quoi parle-t-on ?

Philippe Cudré-Mauroux*

Table des matières

Résumé 3
Zusammenfassung 3
1. Intelligence artificielle : une renaissance 4
2. L'apprentissage profond 4
3. L'entraînement d'un modèle 5
4. Les applications d'apprentissage profond aujourd'hui 7
5. Conclusions 8

Résumé

Le *Big Data* est aujourd'hui le carburant de choix d'une nouvelle révolution, celle de l'intelligence artificielle. La combinaison d'énormes quantités de données et de nouvelles capacités computationnelles a ramené sur le devant de la scène une technologie de près de 40 ans : les réseaux de neurones artificiels. Leur incarnation moderne, l'apprentissage profond, permet d'entraîner automatiquement des modèles qui peuvent aujourd'hui dépasser les performances humaines dans certains cas d'application spécifiques (*e.g.* jeux de stratégie, reconnaissance d'objets). Dans cette contribution, je caractérise cette révolution en marche ainsi que ses conséquences à court terme. J'esquisse notamment son formidable potentiel dans plusieurs domaines, reviens sur ses limitations actuelles, parfois surprenantes, et sur le danger qu'elle porte à notre autonomie et à notre vie privée.

Zusammenfassung

Big Data ist heute der wichtigste Treibstoff einer neuen Revolution, die der künstlichen Intelligenz. Die Kombination aus riesigen Datenmengen und neuen Rechenkapazitäten hat eine fast 40 Jahre alte Technologie in den Vordergrund gerückt: künstliche neuronale Netze. Ihr moderner Ausdruck, *deep learning*, ermöglicht es, Modelle automatisch zu trainieren, die heute

* Professeur ordinaire à l'Université de Fribourg.

in bestimmten Anwendungsfällen (z.B. Strategiespielen, Objekterkennung) die menschliche Leistung übertreffen können. In diesem Beitrag beschreibe ich diese Revolution sowie ihre kurzfristigen Folgen. Insbesondere skizziere ich das enorme Potenzial künstlicher Intelligenz in mehreren Bereichen, überprüfe ihre gegenwärtigen Beschränkungen, die manchmal überraschend sind, und die Gefahr, die sie für unsere Autonomie und Privatsphäre darstellt.

1. Intelligence artificielle : une renaissance

L'intelligence artificielle (IA, ou *AI* en anglais pour *Artificial Intelligence*) n'est pas un domaine de recherche nouveau, même si elle fait beaucoup parler d'elle ces jours. L'IA est un sous-domaine de l'informatique qui date des années 1950 et qui a pour but de créer (ou de simuler) des formes d'intelligence se rapprochant de l'intelligence humaine à partir d'outils computationnels. Historiquement, on distingue l'IA *faible*, qui est une intelligence artificielle limitée à un domaine d'application particulier, de l'IA *forte* qui ressemble à une intelligence humaine dans le sens où elle peut accomplir des tâches diverses, raisonner, et apprendre par elle-même.

A l'heure actuelle, et malgré les sommes colossales investies dans la recherche en IA, nul ne sait comment créer une IA forte et il n'y a pas de consensus des spécialistes quant à son apparition éventuelle (ou non) dans les années ou décades à venir. Les progrès de l'IA faible ont par contre été impressionnants durant ces dernières années. Alors que les techniques d'IA se basaient beaucoup sur les logiques formelles dans les années 1980 et 1990, le renouveau de l'IA est en grande partie dû à des approches différentes, statistiques, venant de *l'apprentissage automatique* (ou *Machine Learning* en anglais). L'apprentissage automatique n'encode pas le savoir ou l'intelligence sous forme logique, mais tente au contraire de développer des intelligences de manière itérative, en utilisant des algorithmes qui vont perfectionner des modèles au fil de leurs expériences ou de leurs confrontations avec des données.

2. L'apprentissage profond

L'apprentissage profond (*Deep Learning* en anglais) est la technique d'apprentissage automatique la plus en vue, à l'heure actuelle, et est en grande partie responsable du regain d'intérêt pour l'IA. L'apprentissage profond se base sur des modèles computationnels appelés réseaux de neurones artificiels. Même s'ils sont inspirés de neurones humains, ces modèles n'ont finalement pas grand-chose à voir avec notre cerveau et sont constitués de fonctions mathématiques simples (*e.g.* des tangentes hyperboliques) qui sont connectées en

cascade sur plusieurs couches (d'où leur qualificatif de réseaux *profonds*). La Figure 1 ci-dessous résume la relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond.

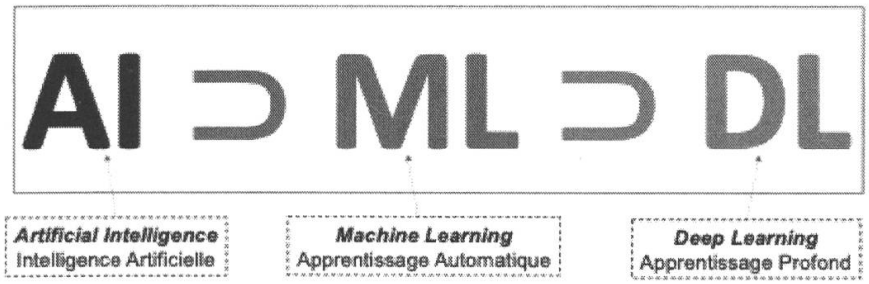

Figure 1 : l'apprentissage profond est un sous-domaine de l'apprentissage automatique, qui lui-même est un sous-domaine de l'intelligence artificielle.

L'apprentissage profond n'est pas nouveau. Yann LeCun, l'un des principaux pionniers du domaine, a publié un algorithme central dans ce contexte (l'algorithme de *rétropropagation du gradient*) en fin 1980 déjà. Cependant, son exploitation fut limitée pendant des années à cause d'un manque de puissance de calcul et d'un manque de données pour entraîner ces modèles (cf. ci-dessous Section 3). Ces contraintes s'estompèrent dans les années 2010, qui virent l'apparition de processeurs graphiques (*GPUs* en anglais pour *Graphics Processing Units*) puissants ainsi que de très grandes quantités de données (*Big Data*). Cette nouvelle puissance de calcul permit d'entraîner de grands réseaux de neurones artificiels (plusieurs centaines de milliers de neurones) sur de très grandes quantités de données (plusieurs millions d'exemples) dès le début des années 2010 déjà, avec des résultats impressionnants pour des tâches répétitives comme la reconnaissance d'images.

Depuis, la taille des réseaux de neurones artificiels a explosé. A titre d'exemple, un modèle récent nommé GPT-3 (un modèle linguistique utilisé pour générer ou analyser des données textuelles) est composé de dizaines de milliards de neurones artificiels, entraînés sur des centaines de milliards de mots.

3. L'entraînement d'un modèle

Comment ces modèles profonds apprennent-ils ? A l'heure actuelle, deux possibilités principales s'offrent aux ingénieurs. L'apprentissage dit supervisé (*supervised* en anglais) se base sur de grandes quantités de données (*Big Data*) qui sont utilisées pour paramétrer les modèles, tandis que l'apprentissage par renforcement (*reinforcement learning*) optimise un agent artificiel sur la base d'expériences répétées, souvent en utilisant un simulateur.

Prenons l'exemple de l'apprentissage supervisé, qui est la technique d'apprentissage la plus populaire à l'heure actuelle. Imaginons une application où un réseau de neurones simplifié (cf. Figure 2) est entraîné à reconnaître des chats sur des images (il s'agit en fait d'un cas courant de *classification binaire*, étant donné qu'il y a uniquement deux sortes de résultats : l'image contient un chat, ou elle n'en contient pas).

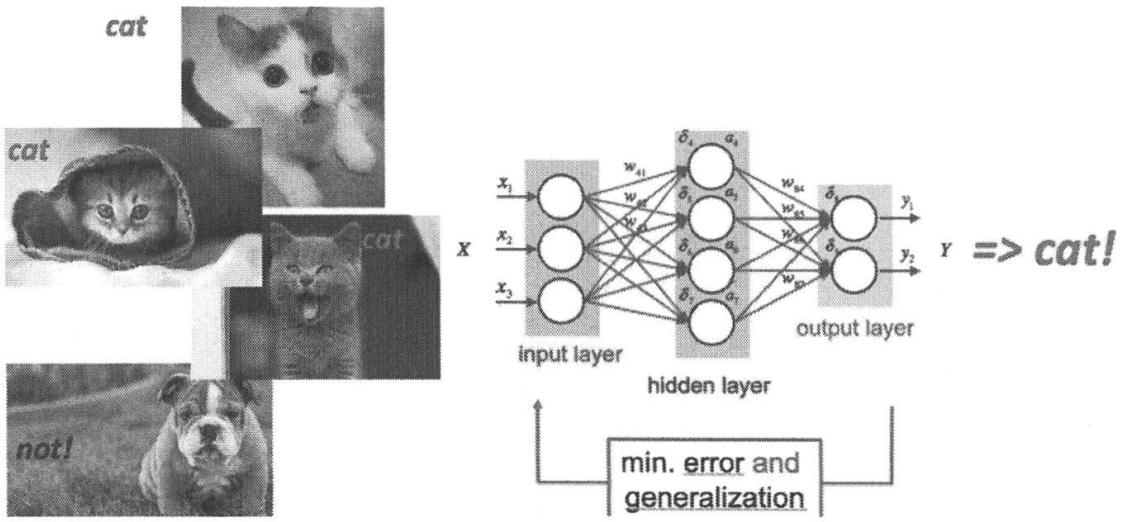

Figure 2 : un réseau (simplifié) de neurones artificiels est entraîné afin de reconnaître des chats sur des images.

L'apprentissage (ou *entraînement* du modèle) s'effectue de manière itérative, en soumettant des exemples d'images les uns après les autres en entrée du réseau de neurones. Le réseau de neurones, dont la topologie (nombre de neurones et connections) est fixe, est en fait un approximateur de fonction, qui calcule en sortie une valeur représentant la probabilité qu'un chat apparaisse sur l'image donnée en entrée. La différence entre la valeur en sortie du réseau et la valeur correcte (*i.e.*, le *label* « chat » ou « pas chat », qui doit être connu pour chaque image d'entraînement) est utilisée de manière itérative pour ajuster les poids du réseau de neurones (*w* sur la Figure 2) et faire converger petit à petit le modèle vers des résultats corrects. Après cette phase d'entraînement, durant laquelle potentiellement des millions d'exemples vont être utilisés pour ajuster les poids du réseau petit à petit, le modèle pourra être utilisé sur des images non labellisées, qui pourront alors être classifiées automatiquement (avec, bien entendu, une certaine marge d'erreur, qui dépendra notamment de la topologie du réseau, de l'algorithme d'apprentissage, ainsi que de la qualité et de la quantité des images d'entraînement).

4. Les applications d'apprentissage profond aujourd'hui

L'apprentissage profond a permis de faire un bond dans de nombreux domaines d'application. Aujourd'hui, les réseaux de neurones sont, par exemple, utilisés pour analyser des images ou des vidéos, pour créer, analyser ou traduire du texte, ou alors pour jouer à différents jeux. Dans tous ces domaines, les réseaux de neurones sont maintenant aussi performants, voire plus performants, que des êtres humains, et sont également beaucoup plus rapides. De manière générale, l'apprentissage profond permet, à l'heure actuelle, d'automatiser de nombreuses tâches répétitives, pour peu qu'on ait suffisamment de données labellisées (ou un environnement de simulation performant) à disposition.

Malgré ces résultats impressionnants, il est actuellement encore difficile d'employer ces modèles dans de nombreux cas d'application sensibles ou requérant des résultats systématiquement fiables. En effet, ces modèles sont souvent qualifiés de *black box*, dans la mesure où ils sont très difficiles à ausculter, débugger ou même visualiser (ils sont en fait constitués de dizaines de millions de composants et paramètres, qui sont quasiment impossibles à analyser pour des êtres humains). De plus, ils sont souvent biaisés (en raison, notamment, des données d'entraînement, qui sont souvent elles-mêmes biaisées) et produisent des résultats aléatoires lorsque les données analysées sont inhabituelles ou trop différentes des données utilisées pour l'entraînement (cf. par exemple la Figure 3, qui montre comment un panneau de signalisation légèrement modifié peut gruger les modèles de reconnaissance d'images les plus performants).

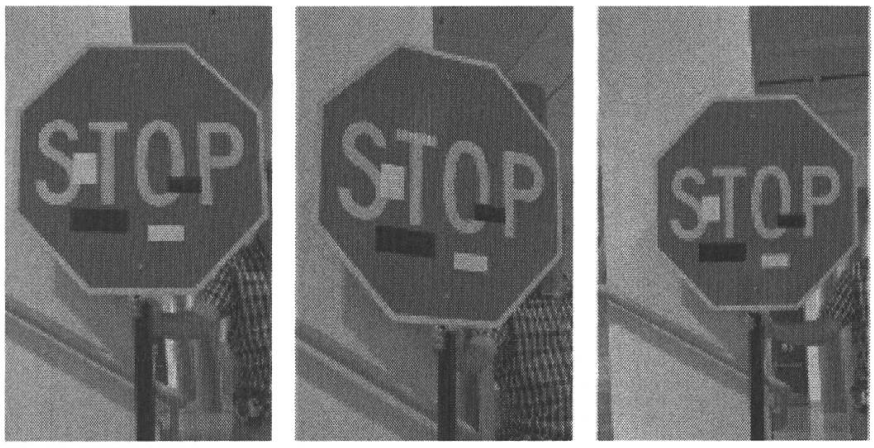

Figure 3 : un panneau de stop légèrement modifié (avec de petits autocollants blancs et noirs) gruge les réseaux de neurones les plus sophistiqués, qui du coup le reconnaissent comme un panneau de limitation de vitesse (45 mph) [K. Eykholt et al. CVPR 2018]

5. Conclusions

L'intelligence artificielle est aujourd'hui à un nouveau carrefour. Après des progrès impressionnants dans les années 2010, où l'apprentissage supervisé de réseaux de neurones profonds a permis des avancées spectaculaires dans l'analyse automatique d'images ou de textes, les défis sont aujourd'hui nombreux.

D'un point de vue technique, plusieurs problèmes fondamentaux subsistent. Citons notamment :

- La dépendance au *Big Data* : actuellement, l'entraînement de réseaux de neurones profonds reste l'apanage des très grandes sociétés d'informatique, qui elles seules ont suffisamment de données labellisées pour entraîner ces réseaux. De nombreux projets visent à limiter cette dépendance, en adaptant des modèles déjà entraînés pour un autre contexte (*transfer learning*) ou en se basant sur des connaissances préalables afin de minimiser le nombre d'exemples d'entraînement (*e.g. few shots learning*).
- La verticalité et la fragilité des modèles : comme expliqué ci-dessus, les intelligences artificielles actuelles apprennent d'exemples labellisés. Leur performance est du coup directement liée à la qualité et à la diversité de ces exemples. Si ces exemples sont biaisés, ce biais va se répercuter sur le modèle dans bien des cas, voire être accentué. De plus, le modèle ne sera vraiment performant que sur des données ressemblant de près ou de loin à ces exemples, avec des résultats souvent catastrophiques (cf. Figure 3) lors de l'analyse de données différentes ou surprenantes.

Au-delà de ces problèmes techniques fondamentaux, qui vont prendre des années, voire des décades, à être résolus, les déploiements actuels d'intelligences artificielles posent de nombreux défis sociaux. Le premier concerne toujours les données ; dans bien des cas, le recours aux données personnelles de tout un chacun est la seule manière d'obtenir un modèle performant. L'écosystème digital intrusif que nous connaissons en découle directement. Les géants du numérique actuels déploient une sophistication technique folle afin de tracer leurs clients et de récupérer leurs données pour finalement optimiser leurs propres processus via des modèles d'apprentissage profond, avec toutes les dérives que cela peut engendrer (manque de transparence, fuite de données, utilisation de ces données privées à des fins idéologiques ou politiques, etc.).

Un dernier point sensible concerne la place de la Suisse, et même de l'Europe, dans ce contexte. Alors que l'intelligence artificielle est en train de bouleverser non seulement notre paysage digital mais également des pans entiers de notre économie (*e.g.* marketing, finance, services en tout genre et même industrie), elle reste largement dominée par des sociétés américaines et chinoises. Les

géants européens ont accumulé un tel retard dans ce domaine extrêmement technique et compétitif, qu'il est à ce jour difficile de voir comment ils pourraient le rattraper. Il en résulte une lente érosion de la puissance économique du vieux continent, avec un nombre croissant de biens et services supplantés par des homologues plus efficaces et optimisés par l'IA. Il en découle également une lente érosion de notre capacité d'autodétermination, alors que de plus en plus de décisions affectant notre quotidien (*e.g.* concernant nos déplacements, achats, lectures ou même nos activités de loisir) sont pilotées par des modèles entraînés loin d'ici, à des milliers de kilomètres de notre pays.