

LE PRINTEMPS DES MACHINES

Olivier Perriquet

Colloque « *Archéologie des Médias, Écologies de l'Attention* »

Centre Culturel International de Cerisy, 30 mai - 6 juin 2016

<http://www.ccic-cerisy.asso.fr/media16.html>

L'intelligence artificielle a subi une série de phases d'hibernation, dont l'une des plus importantes a fait suite aux prédictions pessimistes de Marvin Minsky quant à la capacité des réseaux de neurones artificiels à rivaliser avec des méthodes algorithmiques. Mais les succès récents du paradigme connexionniste semblent augurer du retour d'un printemps ensoleillé pour une telle approche de l'intelligence artificielle, inspirée du vivant. Par ailleurs, l'objet même de l'IA s'est déplacé. Le modèle de la machine intelligente est passé de celui du joueur d'échec, trouvant son paroxysme avec la victoire médiatique de l'ordinateur Deep Blue ou du programme AlphaGo, à celui du petit enfant partant explorer le monde qui l'entoure. La difficulté des jeux combinatoires (les échecs, les grammaires des chatbots, etc.) est en effet bien inférieure à celle qui consiste à acquérir les rudiments d'un langage inconnu en engageant son corps et non ses seules facultés intellectuelles. Il existe à cet endroit un déplacement épistémologique que je me propose d'explorer.

Nous avons assisté récemment à la victoire très médiatisée d'un programme informatique, nommé AlphaGo [Sil16], contre Lee Sedol, joueur de go professionnel coréen, considéré comme l'un des meilleurs au monde. Cette victoire en rappelle une autre, elle aussi spectaculaire et également d'une grande portée symbolique dans le développement d'une intelligence non-humaine, celle de Deep Blue sur Garry Kasparov aux échecs en 1997 [Hsu02]. Le jeu de go résistait depuis une vingtaine d'années aux efforts des chercheurs en intelligence artificielle pour plusieurs raisons, en particulier du fait d'une combinatoire beaucoup plus élevée qu'aux échecs car le goban est très grand, comparé à l'échiquier, et le nombre de coups possible est par conséquent plus important, mais également à cause des principes de jeu et de la façon dont s'élaborent les stratégies au go, où les positions occupées comportent plus d'ambivalence qu'aux échecs car elles participent à une conquête de territoires dont l'intérieur et l'extérieur n'apparaissent que progressivement au cours de la partie, rendant ainsi la définition même de ce qui constitue le contour d'un territoire délicate à établir (le go est un jeu constructif où l'on se partage un espace commun ; les échecs sont un jeu destructif où il s'agit d'anéantir l'adversaire...). Vingt ans après les échecs, le go est donc désormais tombé aux mains des machines. Pour autant la machine ne *pense* pas, serions-nous tentés de nous rassurer, c'est simplement sa puissance de calcul qui l'emporte sur l'intelligence créative dont fait preuve le joueur humain. Tout se passe comme si l'espace symbolique qu'elle est en mesure d'explorer dépassait les capacités cognitives humaines, et c'est sans doute effectivement le cas. Après tout, une simple calculatrice ne nous bat-elle pas à la multiplication ? Bien évidemment, la supériorité d'une intelligence artificielle à un jeu où les humains se mettent mutuellement à l'épreuve, confrontent leur inventivité et démontrent leurs talents stratégiques, suscite nettement plus de trouble. Mais pourquoi cette nouvelle étape est-elle malgré tout si importante, n'avions nous pas déjà franchi ce cap emblématique avec Deep Blue deux décennies plus tôt ?

En réalité, si les deux programmes – AlphaGo (2015) et Deep Blue (1997) – procèdent d'une façon similaire en explorant la combinatoire des coups possibles, leurs approches sont essentiellement différentes. Les deux jeux n'offrent d'ailleurs pas la même résistance. Dans le cas des échecs, les rapports de force s'expriment au travers de la succession des coups qui vont rendre possible ou non l'occupation de telle ou telle position et, en définitive, la prise de la pièce maîtresse de l'adversaire, tandis qu'au go, nous sommes face à un problème d'une difficulté équivalente à celui de la reconnaissance de formes. Alors que Deep Blue cherche à optimiser une fonction de coût qui reflète numériquement l'avantage du joueur, en utilisant un algorithme (Minimax) et une méthode heuristique (Alphabeta) qui sont devenus un classique de la théorie des jeux, le programme AlphaGo utilise des méthodes d'apprentissage par réseaux de neurones artificiels, et notamment des techniques d'apprentissage profond (deep learning) dont les résultats spectaculaires ces dernières années ont

surpris jusqu'aux experts du domaine. Nous avons ainsi d'un côté un algorithme, de l'autre un réseau de neurones, deux méthodes dont les présupposés idéologiques opposés correspondent à des paradigmes distincts en intelligence artificielle.

Le Minimax utilisé par Deep Blue est un *algorithme*. Il détermine un coup optimal en minimisant et maximisant alternativement une fonction d'évaluation selon une procédure mathématique explicite. Le terme « algorithme », qui vient du nom d'un mathématicien persan ayant vécu au 8^{ème} siècle – Muhammad Ibn Mūsā al-Khwarizmī – traduit en latin par Algoritmi, désigne une méthode mathématique qui peut s'énoncer de manière symbolique et univoque et s'appliquer mécaniquement en vue d'obtenir un résultat précis ou de résoudre un problème donné. L'algorithme d'Euclide, qui décrit un procédé automatique de calcul du plus grand diviseur commun entre deux nombres entiers, en est un bon exemple. A l'origine liée aux nombres entiers, la notion d'algorithme s'est étendue par la suite à la manipulation d'objets de plus en plus complexes, tels que du texte, des images, des formules logiques, des structures mathématiques, des objets physiques, etc. Les amateurs de Rubik's cube savent par exemple résoudre le célèbre casse-tête par une série de mouvements qu'ils appliquent automatiquement, et dont l'expression formelle, ressemblant à des énoncés du type « $L^2(ED)(R'L)F^2(RL')(ED)R^2 ; y(R'UR'U)(R'U'R'U)(RUR^2) ; y^2R^2(ED)(RL')B^2(R'L)(ED)L^2$ [...] » rappelle le code des langages de programmation. Le petit nombre d'opérations permises et le caractère systématique de leur enchaînement amène en effet à concevoir un jeu d'écriture qui en simplifie l'expression, et cette graphie compacte et synthétique nous rappelle précisément que derrière la rhétorique ou les manipulations qui les expriment, les opérations en jeu sont bien de nature symbolique.

Un réseau de neurones artificiel est un modèle informatique inspiré du fonctionnement des neurones biologiques, mais dont le principe d'action ne s'exprime pas sous la forme d'un algorithme. Il s'agit d'un artefact construit pour être utilisé comme on se servirait d'un marteau, d'une fourchette ou d'un microscope... Ce sont sa fonction et ses possibilités d'action sur le réel qui le définissent en premier lieu, de la même façon qu'un outil (le terme anglais d'*affordance* traduit assez bien cette idée). Les neurones naturels, dont s'inspirent ces modèles, sont des cellules nerveuses qui transmettent un signal bioélectrique et dont on peut schématiser le fonctionnement ainsi : chaque neurone est connecté en entrée à un certain nombre d'autres neurones, dont il reçoit les influx nerveux, et en sortie à un petit groupe de neurones, à qui il transmet son influx, sous la forme de signaux intermittents. Si le neurone est excité au delà d'un certain seuil par la somme des influx qu'il reçoit, il émet une décharge le long de son axone vers les neurones auquel il est relié. Un neurone « formel », version abstraite et simplifiée du neurone biologique, est défini par la donnée d'un certain nombre d'entrées, d'une unique sortie, et d'une fonction de transfert simple qui détermine sous quelle condition et à partir de quelle

valeur le neurone doit activer sa sortie. Un réseau de neurones artificiel est un assemblage de neurones formels réalisé par le câblage de certaines sorties sur certaines entrées. En pratique, pour construire un réseau de neurones, on fixe d'abord sa topologie, c'est à dire l'agencement des connections, le réseau comporte alors un certain nombre d'entrées et au moins une sortie ainsi que de nombreuses connections internes, pouvant être organisées hiérarchiquement, en niveaux successifs (c'est le cas des réseaux convolutifs du deep learning [Sch15]). Cette étape est largement empirique. On doit ensuite réaliser une phase d'apprentissage, et il faut alors disposer d'un corpus d'exemples qu'on présentera à l'entrée du réseau. Cette étape procède de façon incrémentale, en ajustant à chaque étape les pondérations des influx nerveux dans le réseau. Certaines méthodes d'apprentissage consistent à modifier par exemple l'efficacité des connections de façon à renforcer la connectivité des neurones dont on observe une activation synchrone (« *neurons that fire together, wire together* » [Heb49]). Il s'agit, en définitive, de rendre le réseau apte à une tâche donnée : par exemple la reconnaissance d'un type de caractère (le corpus d'apprentissage est alors constitué par une variété de caractères dont on connaît déjà l'appartenance ou non à la classe à reconnaître), ou bien la classification d'un ensemble d'images en différentes catégories, l'inférence de concepts à partir d'un texte, la détection et le suivi d'objets dans un environnement, etc. Une fois entraîné, le réseau est opérationnel et rigide, et spécialisé dans une tâche donnée : il fonctionne comme un système opaque, comportant des entrées, où on lui présente des données, et des sorties, où il exprime un prédicat.

L'intelligence artificielle s'est organisée au cours de son développement en deux paradigmes, dont on trouve les prémisses dans l'article séminal d'Alan Turing [Tur50] et dont chacun de ces programmes est emblématique : le paradigme « logique » ou « fonctionnaliste » où un problème s'exprime sous forme symbolique et sa résolution sous forme algorithmique ; et le paradigme « biomimétique » ou « connexionniste », où la solution d'un problème passe par la construction et l'entraînement d'un modèle, inspiré du vivant. Nous pouvons dire, au risque d'être un peu schématique, qu'ils ont eu le vent en poupe à tour de rôle. La première approche est liée au développement des sciences cognitives, dont le postulat est qu'il est possible d'isoler un niveau autonome, celui des représentations, indépendamment des processus physiques en jeu et du support sur lequel elles s'inscrivent, tandis que la seconde est issue de la cybernétique, qui s'est toujours refusée à faire cette abstraction. Chacune de ces approches a subi une série d'« hivers », dont l'un des plus importants a fait suite aux prédictions pessimistes de Marvin Minsky quant à la capacité des réseaux de neurones artificiels à rivaliser avec des méthodes algorithmiques [Min69], qui avait précipité à l'époque l'arrêt des recherches connexionnistes pendant dix ans. Or la vogue connexionniste actuelle et les progrès de la vie artificielle pourraient nous conduire aujourd'hui à revenir sur le jugement de l'IA symbolique et à rouvrir le débat philosophique sur la nature de l'information [Tri08].

Si des représentations internes semblent émerger dans les différentes couches des réseaux de neurones convolutifs, et celles-ci présentent d'ailleurs des similarités avec ce qui est observé dans le cortex visuel humain [Kri15], ce sont des motifs visuels aux contours diffus, qui sont tout au mieux *pré-symboliques*, nous emmenant sur le terrain de la reconnaissance de formes plus que de la représentation symbolique. Bien qu'en pratique les réseaux de neurones artificiels soient généralement implémentés sur le même type de hardware que les algorithmes (préférentiellement les processeurs des cartes graphiques, dont l'architecture est optimisée pour la multiplication matricielle), ces modèles font circuler en premier lieu des signaux et non des signes, nous ramenant ainsi très directement à la matérialité au travers de laquelle s'inscrit l'information. Certaines recherches visent d'ailleurs au développement de supports matériels dédiés imitant les tissus biologiques vivants [Dar16] ou utilisant ces derniers [War10] dont la miniaturisation et la faible consommation énergétique sont hors de portée des hardware actuels. Un réseau de neurones est un objet plastique. Les réseaux naturels se développent par des processus de morphogenèse (il existe au cours de leur maturation des moments privilégiés où ils présentent une plasticité plus élevée [Pro89]) et il est tout à fait envisageable de pousser le biomimétisme jusqu'à ce stade en imaginant des réseaux de neurones artificiels, *in vivo* tout autant qu'*in silico*, qui soient sujets à deux sortes de plasticité : une plasticité qu'on pourrait dire de type « software » où, l'architecture du réseau étant préalablement fixée, l'apprentissage transforme la circulation du flux d'information au sein du réseau, et une autre de type « hardware » qui descendrait un étage plus bas et modifierait la topologie du réseau.

Certes ce sont bien des algorithmes qui régissent la propagation des signaux, mais un réseau de neurones est jugé en définitive par son efficacité pratique. La conception des réseaux de neurones relève aujourd'hui à certains égards d'une pratique d'ingénieur plus que de mathématicien, et son opacité constitutive, que les cybernéticiens qualifiaient de « boîte noire », conduit au retour d'une forme d'empirisme. Or de la boîte obscure à l'obscurantisme, il n'y a que le pas de l'idéologie à franchir et le printemps de l'approche connexionniste pourrait très bien devenir... un automne de la pensée symbolique. La physicalité du support de l'information, reprenant aujourd'hui une importance de tout premier plan, comme en écho à un mouvement similaire qui a eu lieu vers la fin des années 80 et qu'on avait nommée la « thèse de l'embodiment » [Bro90], nous rapproche à nouveau des recherches en vie artificielle et en robotique, qui sont les formes plus incarnées de la recherche en intelligence artificielle. Le modèle de la machine intelligente qui était celui du joueur d'échec, trouvant son paroxysme avec la victoire médiatique de l'ordinateur Deep Blue, laisse sa place à celui du petit enfant partant explorer le monde qui l'entoure, vérifiant que le carré rentre bien dans le carré, le rond dans le rond, ou l'étoile dans l'étoile... L'ambition monte alors d'un cran car la difficulté des jeux combinatoires est fort inférieure à celle qui consiste à acquérir les rudiments d'un langage inconnu en engageant son corps et non ses seules facultés intellectuelles. Le jeu de l'enfant, si l'on veut, décline

en complexité celui de l'adulte. Ce retour au corps (perception, reconnaissance de motifs, préhension, voire émotions, apprentissage, importance des formes) préfiguré par les avancées spectaculaires des réseaux de neurones artificiels face à l'IA symbolique, comme refaisant surface, nous engage vraisemblablement vers une cohabitation prochaines avec d'autres entités, munies de facultés comparables aux nôtres et avec lesquelles il nous faudra interagir non plus de main à outil mais... de pair à pair. Et ce constat s'accompagne d'un certain vertige, car si la loi empirique de Moore (celle qui affirme que les capacité de stockage et de calcul doublent tous les deux ans) continue à être vérifiée dans les années à venir, nous aurions théoriquement la capacité matérielle de réaliser un réseau de neurones de la taille de celui d'un cerveau humain vers le milieu du 21ème siècle. On peut prédire qu'à ce stade, une pensée consciente d'elle-même trouvera à s'exprimer dans cet esprit non-humain dont la corporéité nous est aujourd'hui encore inconnue, et nous retrouverions alors, à l'étage le plus haut, la pensée symbolique perdue quelques étages plus bas.

BIBLIOGRAPHIE

- [Heb49] D.O. Hebb, The organization of behavior. New York, Wiley, 1949
- [Bro90] R.A. Brooks, Elephants Don't Play Chess. Robotics and Autonomous Systems 6, 1990
- [Hsu02] F.H. Hsu, Behind Deep Blue: Building the Computer that Defeated the World Chess Champion. Pr Univ Press, 2002
- [Min69] M. Minsky & S. Papert, Perceptrons. An Introduction to Computational Geometry. M.I.T. Press, Cambr Mass, 1969
- [Pro89] A. Prochiantz, La Construction du cerveau. Hachette, 1989
- [Kri15] N. Kriegeskorte, Deep neural networks: a new framework for modelling biological vision, Annu. Rev. Vis. Sci. 2015. 1:417-46.
- [Sch15] J. Schmidhuber, Deep Learning in Neural Networks: An Overview. Neural Networks, Vol 61, Jan 2015, pp 85-117
- [Sil16] D. Silver et al. Mastering the game of Go with deep neural networks and tree search. Nature, Vol 529, 7587, pp 484-489
- [Tri08] M. Triclot, Le moment cybernétique: La constitution de la notion d'information, Champ Vallon, 2008
- [Tur50] A.M. Turing, Computing machinery and intelligence. Mind, 59, 433-460
- [War10] K. Warwick et al. Controlling a Mobile Robot with a Biological Brain. Defence Sc. Journal, Vol 60 n1, pp 5-14, 2010
- [Dar16] DARPA, projet SyNAPSE – <http://www.artificialbrains.com/darpa-synapse-program>