

UN LIVRE DE YANN LE CUN : Quand la machine apprend

Par [JFB](#) le lun 08/01/2024 - 20:33



L'intelligence artificielle, sur le terrain

Yann Le Cun, leader de l'intelligence artificielle chez Facebook, est avec Joseph Sifakis un des deux Français [1] à avoir obtenu le Prix Turing (édition 2018, avec Yoshua Bengio et Geoffrey Hinton), comparable pour l'informatique au Prix Nobel ou à la Médaille Fields.

Né en 1960, Yann Le Cun est dès son enfance passionné par le mystère de l'intelligence humaine et animale, cependant que son père ingénieur stimule sa curiosité pour les questions techniques. Il suit des études d'ingénieur à l'ESIEE Paris, que sa curiosité scientifique l'incite à prolonger par un DEA et par un doctorat en informatique à l'Université Pierre-et-Marie-Curie, sous la direction informelle de Françoise Fogelman (informelle parce qu'elle n'est pas encore habilitée). Son sujet de prédilection est déjà l'apprentissage automatique pour l'intelligence artificielle (IA), dès son travail de thèse il perfectionne l'algorithme de rétro propagation du gradient, une méthode fondamentale de ce domaine.

Au début de son livre Yann Le Cun raconte qu'en ces années 1980 les systèmes de réseaux de neurones artificiels pour l'apprentissage automatique ne sont pas un sujet de recherche très populaire ; de ce fait la communauté qui y travaille n'est pas nombreuse, ce qui facilite les rencontres : « Ma vie professionnelle bascule réellement en février 1985 lors d'un symposium aux Houches, dans les Alpes. Je rencontre là-bas la fine fleur de la recherche internationale qui s'intéresse aux réseaux de neurones ». Il fait ainsi la connaissance de Larry Jackel, chef de

l'Adaptive Systems Research Department [2] aux Bell Labs, puis de Geoffrey Hinton, professeur à l'université de Toronto et futur co-récipiendaire du Prix Turing, dans le laboratoire duquel il effectue un stage post-doctoral, en compagnie de son ami Léon Bottou, autre spécialiste français d'IA. C'est à cette époque qu'il rencontre à Montréal Yoshua Bengio, alors étudiant en master, passionné de réseaux de neurones, futur co-récipiendaire.

Après Toronto Yann Le Cun rejoindra les laboratoires AT&T (les fameux Bell Labs). Mais vient un temps où la recherche y périclité, cependant qu'en 1995 commence une nouvelle traversée du désert pour les réseaux de neurones : les réseaux convolutifs (inventés conjointement par Hinton, Bengio et Le Cun) et l'algorithme de rétro propagation du gradient sont à la base du succès actuel de systèmes tels que ChatGPT, mais les ordinateurs d'il y a trente ans n'étaient pas assez puissants pour les mettre en œuvre facilement, alors personne n'y croyait. Alors ce seront NEC, l'université de New York, enfin Facebook.

Je vous renvoie au livre pour le détail trépidant des péripéties qui mèneront l'auteur à un dîner dans la salle à manger de Mark Zuckerberg, qui le convaincra personnellement de diriger la recherche en IA de Facebook. Vous y croiserez des personnages hors du commun, Vladimir Vapnik, John Denker par exemple. Bref, un récit alerte, pédagogique et bien écrit, une bonne introduction à l'IA.

Modèles connexionnistes de l'apprentissage

Les systèmes artificiels intelligents (SAI) à base de réseaux de neurones (par opposition aux systèmes d'intelligence artificielle symbolique) acquièrent leurs capacités par apprentissage : on leur soumet des problèmes, la réponse est évaluée, ils modifient leurs paramètres pour tenter une nouvelle réponse, et ainsi de suite. Du plus simple au plus complet, ces systèmes d'apprentissage automatique font l'objet des chapitres 3, 4 et 5 du livre.

Les SAI modernes, tels ChatGPT, ont des milliards de paramètres : les neurones artificiels sont en fait des fonctions mathématiques, les paramètres du modèle sont les entrées des neurones. La méthode de base pour l'ajustement des paramètres est la rétro propagation du gradient, au développement de laquelle Yann Le Cun a contribué, et qu'il expose de façon très pédagogique dans son livre. Les notions mathématiques mises à contribution étaient à mon époque étudiées en Mathématiques spéciales, mais quiconque aura passé un baccalauréat scientifique

devrait pouvoir suivre l'exposé donné dans le livre, éventuellement en s'aidant de Wikipédia (matrice jacobienne) et sinon on peut sauter cet exposé, signalé par une typographie particulière, et comprendre néanmoins les chapitres suivants. À partir du chapitre 8 l'exposé devient plus simple, si vraiment vous avez du mal vous pouvez sauter directement du chapitre 3 au chapitre 8 en première lecture. L'auteur fournit aimablement des programmes en Python qui permettent de faire des essais.

Les réseaux convolutifs

Les réseaux convolutifs sont la principale contribution de Yann Le Cun et de ses collègues, qui leur a valu le Prix Turing. Alors que la rétro propagation de gradient dispose d'une formulation mathématique explicite (les dérivées partielles et la matrice jacobienne), les réseaux convolutifs me semblent moins faciles à expliquer. Ils font l'objet du chapitre 6. Ils ne modifient pas fondamentalement les principes de l'apprentissage automatique, ils en accroissent l'efficacité de façon spectaculaire. En effet, l'algorithme de rétro propagation de gradient dans un grand réseau de neurones nécessite une puissance de calcul considérable.

L'idée de réseau convolutif est née de la recherche en vision par ordinateur pour la reconnaissance d'objets dans les images, recherche qui a conduit à se pencher sur les mécanismes de la vision chez les animaux, tels qu'explorés par Hubel et Wiesel (Hubel et Wiesel, quelle vision !), Prix Nobel de médecine 1981 pour leurs travaux sur ce sujet. Le système de vision des mammifères est organisé en plusieurs couches, une première couche de neurones simples dont chacun est associé à une aire limitée, un champ récepteur, cependant qu'une seconde couche de neurones plus complexes combine les sorties de la première couche. Pour le formuler de façon très résumée, l'idée est que deux champs récepteurs adjacents ont de bonnes chances de « voir » à peu près la même chose, et que l'on peut leur appliquer la même fonction de transformation, ce qui repose les neurones, et, si l'on transpose le traitement dans un ordinateur, cela diminue la puissance de calcul nécessaire dans des proportions spectaculaires.

C'est en 2012 que les réseaux convolutifs terrassent leurs rivaux, lors du concours annuel ImageNet Large Scale Visual Recognition Challenge (ILSVRC), par le truchement de Geoffrey Hinton et de ses étudiants de l'Université de Toronto. Ils utilisent « un réseau convolutif de grande taille inspiré de ceux que j'avais conçus, programmé pour tourner sur un GPU [3], une carte destinée aux rendus graphiques, très efficace pour faire tourner les réseaux convolutifs. »

Pourquoi les GPU ?

Les GPU sont destinées à traiter des problèmes graphiques, et ils se révèlent très efficaces pour faire fonctionner les réseaux convolutifs, inspirés de travaux sur la vision des mammifères. Ce n'est pas une simple coïncidence : l'affichage d'une image sur un écran d'ordinateur est constitué de pixels minuscules, et il est hautement probable que deux pixels adjacents correspondent à une même zone de l'objet dont l'image est affichée, ou du moins au même objet ; de ce fait ces deux pixels seront probablement soumis au même traitement informatique, par exemple au même déplacement sur l'écran. L'organisation matérielle adaptée à ce type de traitement est dite Single instruction multiple data (SIMD), et elle est présente dans la plupart des GPU. Et comme on vient de le voir, il en va de même pour les réseaux convolutifs, inspirés du cortex visuel.

D'autre part, les unités arithmétiques des processeurs généralistes utilisent des nombres entiers de 32 ou 64 chiffres binaires (bits) : les traitements graphiques n'ont pas besoin d'une telle précision, 8 ou à la rigueur 16 bits suffisent à définir un pixel en niveaux de gris, à multiplier par 3 pour la couleur. De même les entrées et les sorties des neurones artificiels sont des entiers ou des nombres fractionnaires de peu de chiffres binaires.

Ce sont les progrès de la microélectronique et ses innovations techniques qui ont facilité la percée des réseaux convolutifs à partir de 2012. Mais il faut se reporter au livre pour une vision plus complète.

Une critique acerbe du monde académique français

Ce livre est aussi le récit du parcours intellectuel de son auteur, et par cela il formule dès le second chapitre une critique du dispositif français d'enseignement supérieur et de recherche scientifique d'autant plus radicale qu'elle est implicite.

Yann Le Cun prépare sa thèse au sein du Laboratoire de dynamique des réseaux (LDR), une équipe assez marginale, les chercheurs ont des postes ailleurs, les locaux sont plus ou moins squattés, l'unique ordinateur est de récupération. Les atouts de Le Cun : il a une bourse de son école, dont il peut utiliser les ordinateurs, et il a lu tous les articles sur les réseaux de neurones. Ce qui lui vaut d'être invité à présenter un exposé à un symposium aux Houches (dont Françoise Fogelman est co-organisatrice), devant « la fine fleur de la recherche internationale qui s'intéresse aux réseaux de neurones : physiciens, ingénieurs, mathématiciens, neurobiologistes,

psychologues, et notamment des membres d'un tout nouveau groupe de recherche en réseaux de neurones qui s'est formé aux Bell Labs, un lieu mythique pour la communauté scientifique. Grâce aux liens que je noue aux Houches, je finirai par être embauché dans ce groupe trois ans plus tard. »

Aux Houches il rencontre aussi Terry Sejnowski, le coauteur avec Geoffrey Hinton d'un article sur les machines de Boltzmann [4], et à la suite de cette conversation il est invité à présenter une communication à l'école d'été de 1986 sur les modèles connexionnistes à l'Université Carnegie-Mellon (Pittsburgh, États-Unis). Geoffrey Hinton, impressionné par son exposé, l'invite à rejoindre son équipe de l'université de Toronto. Hinton sera dans son jury de thèse, avec Maurice Milgram, Françoise Fogelman, Jacques Pitrat (un des piliers de la recherche en IA symbolique en France, l'autre courant du domaine) et Bernard Angéniol. Ensuite tout s'enchaîne... au Canada et aux États-Unis.

Résumons : un peu par hasard Yann Le Cun découvre en 1980 dans les actes d'un colloque de Cerisy sur l'inné et l'acquis le perceptron, la première machine apprenante de Frank Rosenblatt. Passionné, il profite des mercredis après-midi où l'ESIEE fait relâche pour écumer les rayons spécialisés de la bibliothèque de l'Inria (Institut national de recherche en informatique et automatique) à Rocquencourt, et il découvre vite que, dans le monde occidental, plus personne ne travaille sur les réseaux de neurones. Il profite des puissants (pour l'époque) ordinateurs de son école pour développer un algorithme d'apprentissage dans un réseau de neurones qu'il baptise HLM (pour Hierarchical Learning Machine). En 1983, diplômé d'ingénieur en poche, il dégote le LDR, cette équipe de marginaux qui travaillent sur des sujets marginaux mais proches des siens, et qui vont lui mettre le pied à l'étrier. Ce n'est pas du tout la procédure standard pour devenir chercheur en France !

Aurait-il obtenu un poste dans une institution académique française ? Très probablement non : issu d'une école où l'on entre sans passer par les classes préparatoires, inventeur de son sujet de thèse, il a immédiatement tapé dans l'œil des gens des Bell Labs et de l'université de Toronto, mais en France il n'a retenu l'attention que du LDR, des gens sans postes, sans moyens et sans pouvoir... On peut rapprocher son cas de celui d'Emmanuelle Charpentier, prix Nobel de chimie 2020 avec Jennifer Doudna pour des travaux qui sont en train de révolutionner les thérapies géniques ; elle a fait toute sa carrière en Autriche, en Suède, en Allemagne, mais l'Institut Pasteur, où elle a fait sa thèse, ne s'est pas intéressé à elle : elle ne sortait que de l'université ! Après le Nobel ils ont tenté de l'attirer, elle

les a gentiment envoyés balader. On peut aussi penser à Gérard Mourou, prix Nobel de physique 2018 (avec la Canadienne Donna Strickland dont il avait dirigé la thèse) pour leurs travaux sur les lasers : son cas était encore plus grave, il n'était même pas allé au lycée, mais au cours complémentaire ! Inutile de dire qu'il n'avait aucune chance dans le système français, il a fait toute sa carrière aux États-Unis, où visiblement on a été très content de lui. Bref, ces gens brillants mais non conformes aux standards des élites intellectuelles auto-proclamées françaises ont tout fait dans des pays plus accueillants, ce qui n'empêche pas la France de s'enorgueillir de leurs succès.

À la fin de son livre, Yann Le Cun compare la situation d'un jeune chercheur français à celle de son homologue américain, puis leurs évolutions respectives ; c'est à pleurer : rémunération trois fois moindre, charge d'enseignement double, accès à la possibilité d'encadrer des thèses et d'avoir un laboratoire bien plus tardif, moyens financiers misérables... Il serait temps de se réveiller !

Laurent Bloch

[1] En fait trois, puisque Yoshua Bengio, s'il vit et travaille au Canada, est né à Paris et a la nationalité française.

[2] <http://yann.lecun.com/ex/group/index.html>

[3] Graphics Processing Unit, processeur graphique.

[4] Les machines de Boltzmann sont une méthode pour l'apprentissage des réseaux de neurones, qui sera abandonnée au profit de la méthode de rétropropagation de gradient.

•
Catégorie

Sciences et recherches