

➤pensaient qu'il n'était pas possible de le relever. L'algorithme d'apprentissage pourrait en effet se retrouver coincé dans un minimum local, d'où il ne pourrait sortir en ajustant les paramètres par petites touches.

Or avec mes collègues, nous avons montré que si le réseau de neurones est assez grand, le problème des minima locaux est fortement réduit. Dans un tel réseau, la plupart des minima locaux correspondent à un niveau d'apprentissage quasi équivalent à celui du minimum global.

Bien que les problèmes théoriques de l'optimisation soient, en principe, résolubles, la construction de grands réseaux comportant plus de deux ou trois couches avait souvent échoué. À partir de 2005, les recherches financées par l'Icra ont commencé à porter leurs fruits et les obstacles ont été surmontés progressivement. En 2006, nous sommes parvenus à entraîner des réseaux neuronaux plus profonds en utilisant une technique qui opérait couche par couche.

Par la suite, en 2011, nous avons trouvé un moyen d'entraîner des réseaux encore plus profonds (avec davantage de couches de neurones virtuels) en modifiant les opérations effectuées par chacun des neurones, ce qui les rapprochait davantage des neurones biologiques dans leur façon de traiter l'information. Nous avons aussi découvert qu'en injectant du bruit aléatoire dans les signaux transmis entre les neurones au cours de l'apprentissage – comme ce qui a cours dans le cerveau –, les réseaux apprenaient mieux à identifier correctement une image ou un son.

Par ailleurs, deux facteurs essentiels ont contribué au succès des techniques d'apprentissage profond. Le premier est l'augmentation d'un facteur 10 de la puissance de calcul des ordinateurs grâce aux processeurs dédiés au traitement d'image, ce qui a permis d'entraîner des réseaux plus grands en un temps raisonnable. Le second facteur est que l'on avait désormais accès à d'énormes bases de données étiquetées, avec lesquelles les algorithmes d'apprentissage ont pu s'exercer à reconnaître un « chat », par exemple.

Une autre raison aux récents succès de l'apprentissage profond est que le réseau est capable d'apprendre à effectuer un ensemble de calculs qui construisent ou analysent pas à pas une image, un son ou un autre type de donnée. Avec une profondeur suffisante, ces dispositifs excellent dans beaucoup de tâches de reconnaissance visuelle ou auditive. Des travaux théoriques et expérimentaux récents ont d'ailleurs montré qu'il est impossible d'effectuer efficacement certaines de ces opérations mathématiques sans un réseau assez profond.

Mais que se passe-t-il au cœur du réseau de neurones profond? Chaque couche transforme son entrée et produit une sortie qui est envoyée à la couche suivante (voir l'encadré

page 45). Les premières couches se concentrent sur les détails aux échelles les plus petites, puis les couches suivantes agrandissent les échelles considérées. Plus les couches sont profondes, plus elles vont représenter des concepts abstraits.

RECONNAÎTRE DES VIDÉOS

Jusqu'à récemment, les réseaux de neurones artificiels se distinguaient en grande partie par leur capacité à effectuer des tâches telles que la reconnaissance de motifs dans des images statiques. Mais d'autres types de réseaux neuronaux affichent des résultats intéressants sur des événements dynamiques. Les « réseaux neuronaux récurrents », par exemple, ont démontré leur capacité à effectuer une séquence de calculs pour traiter la parole ou une vidéo. Les données séquentielles sont constituées d'unités (phonèmes ou mots dans le cas de la parole) qui se suivent. La façon dont les réseaux neuronaux récurrents traitent leurs entrées ressemble un peu au fonctionnement du cerveau. Les signaux qui se propagent parmi les neurones changent constamment à mesure que les entrées sensorielles sont traitées.

Les réseaux récurrents parviennent à prédire quel sera le mot suivant dans une phrase et peuvent ainsi produire une nouvelle séquence de mots, l'un après l'autre. Ils peuvent aussi s'atteler à des tâches plus complexes. Après avoir « lu » tous les mots d'une phrase, le réseau est à même de deviner le sens de la phrase entière. Un réseau récurrent distinct peut alors utiliser le traitement sémantique du premier réseau pour traduire la phrase dans une autre langue.

La recherche sur les réseaux neuronaux récurrents a connu sa propre période de stagnation à la fin des années 1990 et au début des années 2000. Mes propres travaux théoriques suggéraient qu'ils rencontreraient des difficultés à apprendre à récupérer de l'information du passé lointain, à savoir les premiers éléments de la séquence en cours de traitement. Mais certains de ces problèmes ont été résolus grâce à des techniques consistant à stocker l'information pour qu'elle persiste durablement. Ces réseaux neuronaux utilisent la mémoire temporaire (mémoire cache) d'un ordinateur pour traiter des informations multiples et dispersées, comme des idées contenues dans différentes phrases réparties au fil d'un document.

Le retour en force des réseaux neuronaux à apprentissage profond après le long sommeil de l'intelligence artificielle n'est pas uniquement un triomphe technique. C'est aussi une leçon de sociologie des sciences. En particulier, il souligne la nécessité de soutenir des idées qui font fi du *statu quo* technologique et d'encourager la tenue d'un portefeuille de recherches diversifié, laissant de la place à des champs provisoirement passés de mode. ■

BIBLIOGRAPHIE

D. SILVER ET AL., *Mastering the game of Go without human knowledge*, *Nature*, vol. 550, pp. 354-359, 2017.

Y. LECUN ET AL., *Deep learning*, *Nature*, vol. 521, pp. 436-444, 2015.

Y. BENGIO ET AL., *Representation learning: A review and new perspectives*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35(8), pp. 1798-1828, 2013.

ImageNet classification with deep convolutional neural networks, 26th Annual Conference on Neural Information Processing Systems (NIPS 2012), 2012.

SUR LE WEB

Cours de Yann LeCun au Collège de France, 2015-2016: <https://www.college-de-france.fr/site/yann-lecun/index.htm>



« Quand les “data scientists” nous simplifient la vie »

En ce début de XXI^e siècle, la science des données s’imisce tous les jours un peu plus dans notre quotidien. Reconnaissance vocale, moteurs de recherche, publicités ciblées, recommandations d’achats... Elle est bien sûr déjà omniprésente dans nos activités numériques. Mais cette *data science* gagne aussi progressivement tous les secteurs d’activités économiques. Décoller en A380, emprunter sereinement son train quotidien ou bien encore se remettre d’une catastrophe naturelle: voici trois cas très concrets où des algorithmes aussi discrets qu’efficaces sont déjà à la manœuvre. Autant d’applications qui nécessitent les compétences de *data scientists* chevronnés... Un secteur d’activité florissant qui pourrait créer trois millions d’emplois en Europe! Et grâce au nombre et à la qualité de ses formations, la France est bien placée pour ne pas laisser filer sa part du gâteau... Rien que sur l’Hexagone, le recrutement de 130 000 *data scientists* sera nécessaire d’ici à 2020. En ferez-vous partie?

SOMMAIRE

- Décoller à l’heure p. II
- Des sinistrés pris en charge plus vite p. III
- Améliorer l’information temps réel p. IV
- Et si vous deveniez *data scientist*? p. V

Cahier spécial réalisé en partenariat avec

