

Une plongée dans les réseaux de neurones

Une des premières thèses de la Chaire Good In Tech* a entre autres pour objectif d'améliorer la compréhension de l'IA, et en particulier des réseaux de neurones. Les premières approches émergent.



L'objectif global des recherches est «d'étudier les impacts des outils dits éthiques by design sur les utilisateurs», décrit Jean-Marie John-Mathews, doctorant à Sciences Po. Elles se déclinent sur trois axes : prendre en compte les questions d'opacité de l'IA, en particulier pour les réseaux de neurones ; mieux contrôler les biais ; et rendre plus éthiques les approches à base d'IA, «qui n'est pas forcément respectueuse des données privées», souligne notre interlocuteur. Démarrés début 2019, les travaux ont pour l'instant porté sur l'opacité des réseaux de neurones. «Les besoins se déclinent dans tous les domaines. Par exemple, si un algorithme utilisé pour de la maintenance prédictive dans un contexte industriel demande une intervention sur un équipement, comprendre les facteurs sous-jacents à cette préconisation permet aux techniciens d'emmener sur site les pièces ad hoc», illustre notre interlocuteur. Autre exemple, il peut s'agir de justifier une décision de refus de prêt par une banque, une justification à donner pour des raisons réglementaires.

Jean-Marie John-Mathews a classé les différentes approches d'explicabilité existant à ce jour. «Elles se divisent en plusieurs familles, a posteriori ou intrinsèques, locales ou globales», décrit-il. Dans un réseau de neurones, les activations entre les couches s'adaptent au fur et à mesure de l'apprentissage. Ce fonctionnement se traduit par la rétro propagation des gra-

dients - comprendre un affinement des coefficients entre les couches - visant à améliorer les résultats. Exemple de méthode locale pour les réseaux de neurones : «Calculer la valeur du gradient pour chaque image sur chaque pixel permet de définir une carte de saillance des zones qui ont le plus contribué dans la propagation du gradient pendant la phase d'apprentissage. Par exemple, lorsqu'il s'agit de reconnaître un chat dans une image, on peut imaginer que la zone correspondant à la moustache soit mise en évidence par la carte de saillance et donc représente une zone déterminante dans la décision finale de l'algorithme», explique Jean-Marie John-Mathews. Toujours dans la famille locale, des approches sont adaptées à d'autres technologies d'IA, telle la méthode Random Forest... Pour les approches a posteriori, «il est possible de réapprendre une seconde fois à partir des résultats du premier classifieur. Si ce second algorithme est facilement interprétable (régression linéaire, arbre de décision, etc.), on aura une interprétation a posteriori (ou post-hoc) du premier classifieur non interprétable», détaille Jean-Marie John-Mathews. Les approches globales consistent «à récupérer les neurones les plus activés et à simuler une image avec», ajoute-t-il. Des approches complémentaires qui pourraient aider à nous éclairer un peu sur ce qui se passe à l'intérieur de la boîte noire.

PATRICK BRÉBION

*voir page 51

ALLÉGER LES TÂCHES ADMINISTRATIVES

Le CEA-List s'est associé avec DocteGestio dans le cadre du projet DIM-IA. Le but est d'utiliser de l'IA pour lire et codifier les documents issus des consultations hospitalières (7,7 millions d'hospitalisations par an en France) en vue de leur remboursement et de la mise à jour du dossier patient.

DES MATRICES ALÉATOIRES MIEUX COMPRIS

Les matrices sont utilisées, entre autres, pour l'analyse des données. Les matrices aléatoires constituent un domaine versatile avec des applications concrètes, par exemple en traitement du signal. Les chercheurs, Rafał Latała, Ramon van Handel et Pierre Youssef ont amélioré la compréhension de certaines d'entre elles, les matrices gaussiennes.

UN TOUR DE FRANCE POUR L'INNOVATION

Bpifrance et le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation ont lancé à la mi-septembre un «Deeptech Tour». Cette tournée des campus universitaires français vise à mieux connecter les écosystèmes de la création et du financement des start-up avec les porteurs de projet du monde académique.