

Traduction, reconnaissance d'images, etc. : vers une IA à tout faire ?

Certes, le Deep Learning fournit des résultats intéressants en matière de reconnaissance vocale, de classification d'images ou de traduction. Mais, dans chaque cas, le modèle algorithmique doit être pensé en fonction de la tâche et optimisé pour cette dernière. Autrement dit, une IA de reconnaissance d'images ne saura pas s'adapter à des travaux de traduction. Contrairement à un humain, dont les capacités cognitives sont multiples.

D'où l'idée du Massachusetts Institute of Technology (MIT) et de Google de mettre au point un modèle algorithmique capable de fournir de bons résultats dans de multiples domaines. « *En particulier, ce modèle unique est entraîné simultanément sur ImageNet (une base de données d'images, NDLR), sur diverses tâches de traduction, sur du sous-titrage d'images, sur un corpus de données de reconnaissance vocale et sur une tâche d'analyse de l'anglais* », écrivent les chercheurs de Google, dans un [article](#). La mise au point de cette IA multitâche est passée par l'assemblage de différentes briques de base issues des différents domaines ciblés. « *De façon intéressante, même si un bloc n'est pas crucial pour une tâche, nous observons que son ajout ne nuit jamais aux performances et, dans la plupart des cas, l'améliore sur toutes les tâches, relève Google. Nous montrons également que des travaux disposant de moins de données bénéficient nettement de l'entraînement mutualisé sur d'autres tâches.* »

Un modèle unique, 8 jeux de données

Ce n'est certes pas, loin s'en faut, la première fois que des chercheurs étudient les modèles de Deep Learning multitâches. Et on sait déjà que les modèles de traitement du langage ou de traduction automatique bénéficient d'un apprentissage diversifié. Mais avec une nuance selon Google. « *Tous ces modèles sont formés à d'autres tâches du même domaine : les tâches de traduction sont entraînées avec d'autres tâches de traduction, les tâches de vision avec d'autres tâches de vision, les travaux sur la parole avec d'autres discours.* » Mountain View, associé au MIT, estime être allé un cran plus loin, en confrontant ce qu'il appelle son MultiModel à 8 jeux de données diversifiés.

Et les résultats sont encourageants : dans leur article, les chercheurs indiquent être parvenus à une « *bonne performance* ». Plus précisément à un niveau inférieur à l'état de l'art sur chacune des 8 tâches prises isolément, mais supérieur à « *de nombreux modèles spécifiques à une tâche étudiés dans un passé récent* ». Les chercheurs présentent leurs travaux comme une première étape et entendent bien améliorer leur MultiModel.

Architecture multimodale

Pour l'heure, l'intérêt de ces travaux de recherche consiste à montrer qu'il est possible d'assembler en un modèle cohérent les techniques les plus avancées actuellement dans chaque domaine (convolutions séparables en profondeur, réseaux de neurones de très grande taille, réseaux de neurones avec mécanisme d'attention). Et de détecter de premiers phénomènes intéressants issus de cet attelage. Comme le fait que les réseaux de neurones de grande taille et ceux dotés d'un

mécanisme d'attention améliorent « *légèrement* » la performance du modèle dans la reconnaissance d'images. « *La tâche qui a le moins besoin de ces techniques* », notent les chercheurs.

« *La clé du succès provient de la conception d'une architecture multimodale dans laquelle le plus grand nombre possible de paramètres est partagé et de l'utilisation conjointes de techniques de calcul issues de différents domaines, conclut l'équipe de recherche. Nous croyons que cela ouvre un chemin vers des travaux futurs intéressants sur des architectures d'apprentissage en profondeur plus générales, d'autant plus que notre modèle montre l'existence d'un transfert d'apprentissage de tâches bénéficiant d'une grande quantité de données vers celles où les données sont plus limitées.* »

A lire aussi :

[Crédit Mutuel : « non, l'IA Watson n'est pas magique »](#)

[Pourquoi la voiture 100 % autonome n'est pas près de rouler](#)

[Airbus : comment le Deep Learning fait décoller la reconnaissance d'images](#)

Crédit photo : Lightspring-Shutterstock