

Les réseaux de neurones profonds ne savent pas adapter à la nouveauté

ces réseaux de neurones profonds ne savent pas du tout adapter à la nouveauté. Les réseaux neuronaux profonds sont spectaculaires quand on les applique à des situations proches de celles apprises par le système. Mais dès qu'on s'en écarte significativement, les résultats peuvent

(*) Un réseau de neurones convolutionnel est un algorithme dont la structure est inspirée du cortex visuel. Il est organisé en plusieurs couches contenant des opérateurs invariants par translation appelés « convolutions ».

LES DIFFÉRENTS TYPES D'APPRENTISSAGE

L'APPRENTISSAGE SUPERVISÉ Un algorithme est entraîné sur une base de données étiquetées, constituée de dizaines de milliers, voire de millions d'exemples annotés par des experts humains. Pour une tâche de reconnaissance visuelle, chaque image X de la base de données est associée à sa catégorie d'appartenance Y. Grâce à ces nombreux exemples annotés, l'algorithme apprend à établir un lien statistique entre les entrées X (l'image) et les sorties Y (la catégorie). Au terme de cet apprentissage, l'algorithme peut prédire avec précision la sortie Y, pour une nouvelle entrée X qu'il n'a jamais vue.

L'APPRENTISSAGE NON SUPERVISÉ consiste à entraîner un algorithme sur une base de données non étiquetées. Il doit trouver par lui-même des structures inhérentes à ces données afin d'en extraire un maximum de connaissances.

L'APPRENTISSAGE PAR RENFORCEMENT Cette méthode, inspirée du conditionnement de certains animaux, consiste à améliorer la prise de décision d'un système d'apprentissage grâce à des récompenses positives ou négatives. Après chaque décision, le système reçoit de son environnement une récompense et apprend à ajuster ses décisions futures de façon à maximiser le nombre de récompenses positives espérées.

être absurdes. « Pour des tâches cognitives comme la reconnaissance visuelle, un algorithme peut ainsi donner des résultats complètement différents de ceux des humains si l'image qu'on lui fournit est trop éloignée de celles apprises lors de sa phase d'entraînement », reconnaît Emmanuel Dupoux. « Nous avons en effet une aptitude à résoudre des problèmes nouveaux, car nous avons une plasticité cérébrale qui n'existe pas chez la machine », ajoute Pierre-Marc Jodoin. Notre cerveau évolue avec le temps au fil de nos apprentissages. La structure de l'algorithme, elle, ne change pas. Seuls ses paramètres varient. » En outre, ces algorithmes ont besoin de très nombreuses données pour être efficaces, là où le cerveau humain est très économe. Un système de reconnaissance de la parole comme celui de nos smartphones apprend à partir d'enregistrements vocaux de dizaines de milliers d'heures et de textes de plus d'un milliard de mots. En comparaison, un enfant apprend à parler en utilisant deux à vingt fois moins d'heures de parole, et sans texte.

Par ailleurs, les algorithmes d'apprentissage profond peuvent être dupés facilement, comme l'a montré l'équipe de Ian Goodfellow, chercheur en intelligence artificielle à l'université de Montréal. Après avoir changé quelques propriétés dans des images, des clichés qui apparaissent à l'œil humain comme un chien devenaient des autruches pour la machine (4) ! « Le problème est tellement grave et généralisé qu'il pose de sérieux problèmes de sécurité informatique sur les technologies utilisant l'apprentissage profond », signale

Emmanuel Dupoux. Il suffit s'en convaincre d'imaginer des dégâts que causeraient de telles manipulations sur des systèmes de pilotage de voitures autonomes ou de diagnostics médicaux... Pour éviter ce genre d'écueil, les chercheurs essaient donc de limiter à jour le fonctionnement interne des réseaux de neurones profonds. Ils ne connaissent que dans les grandes lignes leur façon de traiter l'information, inspirée du fonctionnement visuel des mammifères. Ces réseaux neuronaux sont en fait des entités simplifiées qui reproduisent de façon très simplifiée l'action d'un réseau biologique. Ils se composent de millions et sont organisés en plusieurs couches, plus ou moins nombreuses, définissant la profondeur du réseau.

Poids synaptique

Chaque neurone est doté de plusieurs entrées et d'une sortie. Lesquelles transite l'information numérique. Les entrées des neurones de la première couche reçoivent des informations mentales des données d'entrée : une image, ce sont des triplets de nombres réels correspondant à des couleurs (rouges, vertes ou bleues des pixels). La sortie des neurones est quant à elle connectée aux entrées des neurones de la seconde couche, eux-mêmes connectés à une troisième couche, et ainsi de suite. La connexion entre les neurones est définie par un nombre réel correspondant au « poids synaptique ». Celui-ci définit la force de la connexion. Selon la valeur du poids, les signaux envoyés par deux neurones sont soit amplifiés, soit atténués.

Le cheminement de l'information traverse le réseau, et donc la décision effectuée par le réseau, dépend de la valeur de ces poids synaptiques. Chaque couche du