

> d'autres types de courriels, vous remarquerez que seuls les spams tendent à présenter plusieurs de ces attributs à la fois: par exemple, le Nigeria et la promesse d'un gain de un million de dollars, voilà qui est louche. En fait, des motifs de niveau supérieur très subtils pourraient faire la différence entre les messages indésirables et les utiles: des fautes d'orthographe et des adresses IP non conformes, par exemple. Si vous pouviez les détecter, vous pourriez filtrer les messages indésirables avec précision, sans crainte d'écarter à tort la notification d'expédition de notre Viagra.



L'apprentissage ascendant peut dénicher les indices pertinents pour résoudre ce genre de tâche. Pour ce faire, un réseau neuronal doit passer en revue des millions d'exemples tirés de gigantesques bases de données, tous étiquetés soit «courriel indésirable», soit «message légitime». L'ordinateur extrait alors un ensemble de caractéristiques qui permettent de faire le tri entre les spams et le reste.

APPRENDRE À PARTIR DE MILLIONS D'EXEMPLES

De la même manière, le réseau pourrait inspecter des images tirées d'Internet et étiquetées «chat», «maison», «stégosaure», etc. En extrayant les caractéristiques communes de chaque ensemble d'images (le motif qui distingue tous les chats de tous les chiens), il peut identifier de nouvelles images de chat, même s'il n'a jamais vu ces images-là auparavant.

Une méthode ascendante, nommée apprentissage non supervisé, en est toujours à ses premiers balbutiements, mais elle permet de détecter des motifs dans des données sans aucune étiquette. Elle recherche simplement des groupes de propriétés qui identifient un objet; des yeux et un nez, par exemple, vont

toujours de pair pour former un visage et différent des arbres et des montagnes de l'arrière-plan. L'identification d'un objet dans ces réseaux d'apprentissage profond avancés se fait par une division du travail dans laquelle les tâches de reconnaissance sont réparties entre différentes couches du réseau.

Un article paru en 2015 dans *Nature* montrait jusqu'où vont aujourd'hui les méthodes ascendantes. Des chercheurs de DeepMind, une filiale de Google, combinaient deux techniques ascendantes différentes (l'apprentissage profond et l'apprentissage «par renforcement») d'une manière qui permettait à un ordinateur de maîtriser les jeux vidéo pour console Atari 2600. Initialement, l'ordinateur ne connaissait rien du fonctionnement des jeux. Au début, il faisait des hypothèses aléatoires sur les meilleurs coups possible tout en recevant un retour permanent sur sa performance. L'apprentissage profond aidait le système à identifier les caractéristiques sur l'écran, et l'apprentissage par renforcement le récompensait pour un score élevé. L'ordinateur a atteint une bonne maîtrise sur plusieurs jeux, et dans certains cas il était plus performant que des joueurs humains experts. Cela dit, il échouait complètement sur d'autres jeux qui sont tout aussi faciles à maîtriser pour les humains.

La capacité à apprendre à partir de vastes jeux de données (des millions d'images Instagram, de courriels ou d'enregistrements vocaux) permet d'apporter des solutions à des problèmes qui semblaient auparavant insurmontables, tels que la reconnaissance vocale ou d'images. Cependant, il ne faut pas perdre de vue que mon petit-fils n'a aucune difficulté à reconnaître un animal ou à répondre à une demande orale, même avec des données et un entraînement beaucoup plus limités. Des problèmes simples pour un enfant de 5 ans restent extrêmement difficiles pour des ordinateurs, et beaucoup plus compliqués qu'apprendre à jouer aux échecs.

Les ordinateurs qui apprennent à reconnaître une tête à poils et à moustaches ont souvent besoin de millions d'exemples pour catégoriser les objets que nous savons classer à partir de quelques exemplaires. Après un entraînement intensif, l'ordinateur peut être capable d'identifier une image de chat qu'il n'a jamais vue auparavant. Mais il le fait de manière très différente des généralisations humaines. Parce que le logiciel raisonne différemment, il se produit des ratés. Certaines images de chat ne seront pas étiquetées comme étant des chats. Et l'ordinateur peut à tort classer une image dans la catégorie «chat» même s'il ne s'agit que d'une tache floue aléatoire, qui ne tromperait pas un instant un observateur humain.

L'autre approche de l'apprentissage qui a récemment transformé l'intelligence artificielle fonctionne dans le sens inverse, du haut vers le