bas. Elle part du principe que, si l'on peut obtenir une connaissance abstraite à partir de données concrètes, c'est parce que l'on sait déjà beaucoup de choses et, en particulier, parce que le cerveau comprend déjà des concepts abstraits fondamentaux. Comme des scientifiques, nous pouvons utiliser ces concepts pour formuler des hypothèses sur le monde et faire des prédictions sur l'allure que devraient avoir les données (les événements) si ces hypothèses sont vérifiées – le contraire d'essayer d'extraire des motifs à partir des données brutes elles-mêmes, comme on le fait dans l'approche ascendante.

APPROCHE DESCENDANTE

On peut illustrer parfaitement cette idée en revenant sur la plaie des spams avec un cas réel auquel j'ai été confrontée. J'ai reçu un courriel de l'éditeur d'une revue au nom étrange, faisant référence spécifiquement à l'un de mes articles et me proposant d'écrire un article pour le périodique en question. Pas de Nigeria, pas de Viagra, pas de million de dollars, le courriel ne présentait aucun des indices trahissant généralement un spam. Mais en utilisant ce que je savais déjà et en réfléchissant de manière abstraite au processus qui produit les messages indésirables, j'étais en mesure de juger suspect ce message.

Pour commencer, je sais que les spammeurs tentent d'extorquer de l'argent aux gens en faisant appel à la cupidité humaine, et les universitaires peuvent être aussi avides de publier que l'individu lambda sera avide de gagner un million de dollars ou d'améliorer ses performances sexuelles. Je sais aussi que des revues légitimes «en accès libre» ont commencé à couvrir leurs coûts en faisant payer les auteurs plutôt que les abonnés. De plus, mes travaux n'ont rien à voir avec le titre de la revue. En regroupant toutes ces informations, j'ai émis l'hypothèse plausible selon laquelle il s'agissait d'un spam visant à convaincre des universitaires de payer pour «publier» un article dans une fausse revue. J'ai pu tirer cette conclusion à partir d'un seul exemple, et j'ai pu ensuite tester mon hypothèse en vérifiant l'authenticité de l'éditeur par une requête sur un moteur de recherche.

Un informaticien qualifierait mon raisonnement de «modèle génératif», c'est-à-dire un modèle capable de représenter des concepts abstraits. Ce même modèle peut également décrire le processus utilisé pour produire une hypothèse, le raisonnement qui a conduit à la conclusion que le message pouvait être un simple spam. Ce modèle me permet d'expliquer comment fonctionne cette forme de courriel indésirable, mais il me permet également d'imaginer d'autres formes de spams, ou même un type différent de tous ceux que j'ai déjà vus ou dont j'ai entendu parler. Quand je reçois le message de la revue, le modèle me permet de travailler à rebours, de retracer pas à pas pourquoi cela doit être du spam.

Les modèles génératifs occupaient une place primordiale dans la première vague de l'intelligence artificielle et des sciences cognitives, dans les années 1950 et 1960. Mais ils avaient aussi leurs limites. Tout d'abord, la plupart des motifs discriminants pourraient, en principe, être expliqués par de nombreuses hypothèses différentes. Dans mon cas, cela pourrait être que le courriel était vraiment légitime, même si cela semble improbable. Ainsi, les modèles génératifs doivent incorporer des notions probabilistes, ce qui est l'un des développements récents les plus importants pour ces méthodes. En second lieu, on ne sait pas bien d'où viennent les concepts fondamentaux qui charpentent les modèles génératifs. Des penseurs tels que Descartes et Noam Chomsky ont suggéré que nous naissons avec, mais vient-on vraiment au monde en sachant que la cupidité mène aux escroqueries?

Les modèles bayésiens (l'un des principaux exemples récents de méthode descendante) tentent de traiter les deux questions. Portant le nom du statisticien et philosophe anglais du XVIIIe siècle Thomas Bayes, ils intègrent la théorie des probabilités aux modèles génératifs à l'aide d'une technique appelée inférence bayésienne. Un modèle génératif probabiliste peut vous indiquer la probabilité de voir un motif spécifique dans les données si une hypothèse particulière est vraie. Si le courriel est un message indésirable, il s'adresse probablement à la cupidité du lecteur. Mais bien sûr, un message pourrait aussi éveiller la cupidité sans être indésirable. Un modèle bayésien combine la connaissance que vous avez déjà sur les hypothèses vraisemblables avec les données dont vous disposez pour vous permettre de calculer, avec une assez bonne précision, la probabilité que le courriel soit légitime ou indésirable.

Cette méthode descendante correspond mieux que son homologue ascendante à ce que nous savons des mécanismes d'apprentissage >

Les modèles bayésiens sont des modèles d'apprentissage qui calculent des probabilités