



ANTOINETTE MOREAU DUGAULT POUR SCIENCES ET AVENIR - LA RECHERCHE

DUGAULT

Des algorithmes s'attaquent aux odeurs

Contrairement à la vision ou à l'audition, le sens de l'odorat est très difficilement simulé par l'intelligence artificielle, en raison de la nature des données. Mais une méthode, dite des réseaux de neurones en graphes, est en passe de changer la donne.

Les performances impressionnantes de grands modèles de langage peuvent entretenir l'illusion que l'IA à base d'apprentissage automatique est capable de tout faire. Or, elle bute sur nombre d'impasses. La simulation du sens de l'odorat en est une. D'abord parce que le fonctionnement des récepteurs olfactifs n'est pas encore complètement compris. « Dès avant

6054

molécules associées à des senteurs figurent dans la base de données Leffingwell, la plus vaste du genre. À comparer aux 14,2 millions d'images du corpus ImageNet.

la percée de la vision par ordinateur ou la modélisation du son, on connaissait beaucoup de choses sur les ondes sonores et lumineuses et on savait les mesurer avec précision. Ce n'est pas le cas en matière d'olfaction, les parfumeurs se servent seulement de mots pour décrire les différences entre odeurs », explique Matej Hladiš, doctorant à l'Institut de chimie de Nice (Côte-d'Azur). S'y ajoute la difficulté à

constituer des bases de données d'entraînement. On peut facilement collecter et stocker des images et des sons depuis Internet. Pour les odeurs, c'est plus laborieux. « Il faut faire venir des gens dans un endroit dédié, équipé et ventilé correctement, pour leur présenter des fioles d'odorants », continue Matej Hladiš. D'où des corpus bien moins volumineux que pour les images.



M. HLADIŠ

« Dès avant la percée de la vision par ordinateur, on savait beaucoup de choses sur les ondes lumineuses. Ce n'est pas le cas en matière d'olfaction »

Matej Hladiš, Institut de chimie de Nice (CNRS/université Côte d'Azur)

Néanmoins, début 2023, le chercheur a cosigné une étude montrant de nets progrès dus à une méthode particulière de l'apprentissage automatique : les réseaux de neurones en graphes (GNN). Fin août 2023, des chercheurs d'une start-up issue de Google Research, Osmo, et du centre Monell sur la chimie des sens (États-Unis) ont présenté dans *Science* des résultats issus de la même approche. Une donnée en graphe consiste en plusieurs variables (des nœuds) connectées entre elles par des liens (lire l'encadré). Or, avec ses atomes et leurs liaisons chimiques, une molécule est une donnée particulièrement adaptée à un GNN.

L'odorat repose sur 400 récepteurs olfactifs, et chaque odorant (une molécule correspondant à une odeur) en active une combinaison (le code combinatoire des odeurs). Il existe 2,4 millions de paires de récepteurs olfactifs-odorants, mais seule-

POUR EN SAVOIR PLUS

Sur le site du CNRS, les travaux de l'Institut de chimie de Nice sur les odeurs : sciav.fr/926odeurs

ment 46 700 ont été identifiées et décrites en laboratoire. L'Institut de chimie de Nice a donc entraîné un GNN avec 45 000 de ces paires, puis l'a testé avec le reste. Résultat ? Non seulement le modèle prédit mieux que d'autres approches quel odorant active quels récepteurs, mais il peut extrapoler et prédire le code combinatoire à partir de n'importe quelle molécule. « *Le modèle est encore limité par le volume et la qualité du jeu d'entraînement, prévient Matej Hladiš. Ses résultats sont donc meilleurs sur les récepteurs olfactifs les plus étudiés, mais dès qu'il y aura de nouvelles expériences de laboratoire, il gagnera en précision.* »

La numérisation des odeurs reste un casse-tête

Le projet d'Osmo et du centre Monell vise plus directement à décrire une odeur perçue. L'équipe a puisé dans deux bases professionnelles 5000 molécules associées à des termes comme

« acacia », « balsamique », « cas-sis », « tropical », « café brûlé », etc. pour entraîner un GNN. En parallèle, ont été sélectionnées 15 personnes capables de qualifier correctement un odorant, à l'aide d'un lexique de 55 mots à noter de 1 à 5. Puis humains et algorithme se sont vu soumettre les mêmes 323 odorants et leurs résultats ont été comparés.

Pour un odorant donné, le GNN prédit mieux les termes descriptifs fournis en moyenne par les participants que chacun des participants eux-mêmes. Le modèle est aussi meilleur pour décrire 53 % des molécules que l'humain « médian » (celui qui est moins bon qu'une moitié du groupe mais meilleur que l'autre moitié). Ce genre de technologie pourrait par exemple servir à des parfumeurs qui auraient besoin de remplacer un composant tout en gardant la même senteur. « *Une IA pourrait facilement trouver les bons odorants parmi des milliers dans une liste, épargnant la tâche fastidieuse d'avoir à parcourir des catalogues* » imagine le chercheur Michael Schmucker, qui travaille sur des sujets similaires à l'université du Hertfordshire (Royaume-Uni). Mais la nature insaisissable des odeurs fait de leur numérisation un casse-tête encore loin d'être complètement résolu. « *Même si tous les composants chimiques d'une odeur sont identifiés, il reste très difficile de les combiner aux bonnes proportions.* » ■

Arnaud Devillard

LE MOT DU MOIS

GNN

Les articles de recherche sur les réseaux de neurones en graphes (GNN, pour graph neural networks) ont émergé au milieu des années 2000. Mais la méthode a vraiment décollé vers 2015-2016. Particulièrement approprié à des travaux sur les molécules (médicaments, matériaux, parfums), un graphe peut en fait représenter toute donnée consistant en plusieurs variables reliées. « *Par exemple, les nœuds peuvent être des villes et les liens peuvent montrer si des routes les connectent. Un GNN peut être entraîné sur des données présentées sous forme de graphe et apprendre à prédire d'autres propriétés de graphes* », explique Alvaro Sanchez, chez Google. Ce chercheur a participé à l'élaboration de GraphCast, un GNN de prévision météo présenté fin 2023 (lire S. et A n° 923, p. 18). Dans ce projet, toute la surface de la Terre est représentée sous forme de graphe.