



BA ou DA ? Décoder les syllabes pour montrer les limites de l'intelligence artificielle

Des chercheurs de l'UNIGE démontrent aujourd'hui les limites du décodage de l'activité cérébrale par les techniques modernes de « machine-learning », notamment dans le traitement de la parole.



© UNIGE

Anne-Lise Giraud, professeure au Département des neurosciences fondamentales de la Faculté de médecine de l'UNIGE.

Depuis une dizaine d'années, les chercheurs utilisent des techniques d'intelligence artificielle appelées machine-learning pour décoder l'activité du cerveau humain. Ces algorithmes appliqués aux données de neuroimagerie peuvent reconstituer ce que l'on voit, ce que l'on entend, voire ce que l'on pense. Ils montrent par exemple que les mots dont le sens est proche sont regroupés par zones dans différents endroits de notre cerveau. Pourtant, en enregistrant l'activité cérébrale pendant une tâche simple consistant à dire si l'on entend BA ou DA, des neuroscientifiques de l'Université de Genève (UNIGE) et de l'Ecole normale supérieure (ENS) à Paris démontrent aujourd'hui que les zones cérébrales identifiées par le machine-learning ne sont pas forcément utilisées par notre cerveau pour réaliser une tâche, mais reflètent surtout les associations mentales liées à cette tâche. Si le machine-learning est donc efficace pour décoder l'activité mentale, il ne l'est pas nécessairement pour comprendre les processus précis du traitement de l'information dans le cerveau. Des résultats à lire dans la revue *PNAS*.

Les techniques d'analyse modernes des données neuroscientifiques ont récemment mis l'accent sur l'organisation spatiale de la représentation des sons de la parole par le cerveau, que les chercheurs ont pu cartographier précisément par zone d'activité. Des neuroscientifiques de l'UNIGE se sont alors demandés comment ces cartes spatiales étaient utilisées par le cerveau lui-même lorsqu'il exécute des tâches précises. « Nous avons employé toutes les techniques de neuroimagerie humaine à notre disposition pour tenter de répondre à cette question », explique Anne-Lise Giraud, professeure au Département des neurosciences fondamentales de la Faculté de médecine de l'UNIGE.

Une zone focale pour sélectionner l'information

Les neuroscientifiques de l'UNIGE ont fait écouter à une cinquantaine de personnes un continuum de syllabes allant de BA à DA, les phonèmes centraux étant très ambigus et difficiles à distinguer entre les deux possibilités. Ils ont ensuite examiné à l'aide de l'IRM fonctionnelle et de la magnétoencéphalographie comment le cerveau se comporte quand le stimulus acoustique est très clair, ou au contraire lorsqu'il est ambigu et nécessite une représentation mentale active du phonème et son interprétation par le cerveau. « Nous avons constaté que quelle que soit la difficulté à classer la syllabe entendue entre BA et DA, la décision implique toujours une petite zone du lobe temporal supérieur postérieur », relève Anne-Lise Giraud.

Les neuroscientifiques ont ensuite vérifié leurs résultats sur un patient porteur d'une lésion de la zone précise du lobe temporal supé-

rieur postérieur qui sert à distinguer BA et DA. «Et effectivement, bien que ne souffrant d'aucun symptôme apparent, celui-ci n'était plus capable de distinguer les phonèmes BA et DA, confirmant l'importance de cette petite zone pour le traitement de ce type d'information phonémique», ajoute Sophie Bouton, chercheuse au sein de l'équipe de Anne-Lise Giraud.

Les « faux positifs » du décodage par machine-learning

Mais l'information sur l'identité de la syllabe est-elle juste présente localement, comme le démontre l'expérience des scientifiques genevois, ou plus largement dans notre cerveau, comme le suggèrent les cartes réalisées par machine-learning ? Pour répondre à cette question, les neuroscientifiques ont reproduit la tâche BA/DA chez des personnes qui, pour des raisons médicales, sont porteuses d'électrodes implantées directement dans leur cerveau. Cette technique permet de recueillir l'activité neurale très précisément. Une analyse dite univariée a permis de regarder électrode par électrode et contact par contact quelle zone du cerveau était recrutée pendant la tâche. Ici, seuls les contacts dans le lobe temporal supérieur postérieur étaient actifs, confirmant le résultat de l'étude genevoise.

Cependant, lorsque l'ensemble des données a été soumis à un algorithme de machine-learning permettant un décodage dit multivarié des données, des résultats positifs ont pu être observés dans la totalité du lobe temporal et même au delà. «Les algorithmes d'apprentissage sont intelligents mais ignorants», précise Anne-Lise Giraud. «Ils sont très sensibles et utilisent toute l'information contenue dans les signaux, mais ils ne nous permettent pas de savoir si cette information a servi pour effectuer la tâche ou si elle reflète les conséquences de cette tâche, à savoir la diffusion de l'information dans notre cerveau», continue Valérien Chambon, chercheur au Département d'Etudes Cognitives de l'ENS. Les zones cartographiées en dehors du lobe temporal supérieur postérieur sont donc en quelque sorte des faux positifs. Elles contiennent de l'information concernant la décision prise par le sujet (BA ou DA), mais n'ont pas été mobilisées pour réaliser la tâche.

Ces recherches offrent l'opportunité de mieux comprendre comment notre cerveau se représente les syllabes et invitent à une réflexion salutaire sur l'interprétation des données produites par les algorithmes de « machine-learning » en montrant les limites de l'intelligence artificielle dans certains contextes de recherche.

contact

Anne-Lise Giraud

+41 22 379 55 47

Anne-Lise.Giraud@unige.ch

UNIVERSITÉ DE GENÈVE
Service de communication

24 rue du Général-Dufour
CH-1211 Genève 4

Tél. +41 22 379 77 17

media@unige.ch

www.unige.ch