

➤ milliers d'images. Et l'efficacité est tout de suite vérifiable, car tout le monde sait à quoi ressemble un chat.

Toutefois, le manque de données anthropologiques et paléontologiques pertinentes oblige les chercheurs à faire preuve d'inventivité et à créer leurs propres données. « Nous avons un peu triché, admet Oscar Lao. Nous pouvions utiliser une quantité infinie de données pour entraîner le système d'apprentissage, parce que nous utilisions des simulations. »

Les chercheurs ont imaginé des dizaines de milliers d'histoires évolutives, reposant sur des combinaisons différentes de détails démographiques : le nombre de populations d'humains ancestraux, leur taille, les dates de divergence, leur taux de métissage... À partir de ces histoires simulées, ils ont généré les nombreux génomes des individus modernes qui pourraient en résulter, ces génomes ayant nourri leur algorithme d'apprentissage. De la sorte, le programme a appris quels modèles évolutifs sont les plus susceptibles de produire des génomes virtuels.

L'équipe a ensuite laissé le champ libre à l'intelligence artificielle pour qu'elle identifie les histoires correspondant le plus aux véritables données génomiques. Finalement, le système a conclu qu'un groupe humain jusqu'alors inconnu avait contribué à l'évolution des peuples asiatiques. Ces humains constituaient soit une population distincte résultant du métissage des Denisoviens et des Néandertaliens il y a 300 000 ans, soit un groupe descendant de la lignée dénisovienne peu de temps après cette époque.

Ce n'est pas la première fois que l'apprentissage profond est utilisé pour élucider des questions liées à l'évolution. En 2016, un groupe de recherche, dirigé par Andrew Kern de l'université de l'Oregon, aux États-Unis, a utilisé une approche basée sur des simulations et des techniques d'apprentissage profond pour différencier plusieurs modèles d'évolution des espèces, comme les humains. Ils ont découvert que la plupart des adaptations favorisées par l'évolution ne reposent pas sur l'émergence de nouvelles mutations bénéfiques, mais plutôt sur l'expansion de variants génétiques préexistants.

UN NOUVEL OUTIL ENTRE ESPOIR ET ENGOUEMENT

Bien entendu, cette méthode appelle à la prudence. En effet, si l'histoire évolutive humaine ne ressemble pas aux modèles simulés par lesquels ces systèmes d'apprentissage profond sont entraînés, les résultats produits seront incorrects. Ce problème, Andrew Kern et d'autres tentent de le surmonter.

« Je pense que les applications de l'intelligence artificielle en génomiques sont survenues, concède Joshua Akey, chercheur en

écologie et biologie évolutive à l'université de Princeton, aux États-Unis. L'apprentissage profond est un nouvel outil formidable, mais il s'agit simplement d'une autre méthode, et elle ne va pas résoudre tous les problèmes ni nous révéler tout ce que nous souhaitons savoir sur l'évolution humaine. » D'ailleurs, le même Joshua Akey a utilisé d'autres méthodes pour repérer les fantômes de notre passé (voir l'encadré page précédente).

L'apprentissage profond va donner une impulsion nouvelle à la génétique des populations

Certains experts sont encore plus sceptiques. Toutefois, d'autres paléontologues et généticiens y voient un progrès, une technique utile pour des prédictions quant aux futures découvertes de fossiles et aux variations génétiques qui devraient avoir existé au sein des populations humaines il y a des milliers d'années. « Je pense que l'apprentissage profond va donner une impulsion nouvelle au domaine de la génétique des populations », estime Oscar Lao.

Il pourrait en être de même pour d'autres domaines pour lesquels nous avons des données, mais pas d'idées sur les processus qui les ont produits. En 2016, des physiciens ont eux aussi développé un système d'intelligence artificielle pour filtrer les quantités astronomiques de données produites par le LHC. La recherche en géologie et les méthodes de prédiction des séismes commencent également à bénéficier de ces approches d'apprentissage profond.

Pour Nick Patterson, de l'institut de technologie du Massachusetts, et à l'université Harvard, à Cambridge, aux États-Unis, « nous ne savons pas vraiment où cela nous mènera, nous verrons bien. Mais il est toujours bon de voir naître de nouvelles méthodes. Nous utiliserons toutes les approches disponibles si elles semblent pertinentes. » ■

BIBLIOGRAPHIE

M. MONDAL ET AL., Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania, *Nature Communications*, vol. 10, art. 246, 2019.

S. BROWNING ET AL., Analysis of human sequence data reveals two pulses of archaic Denisovan admixture, *Cell*, vol. 173, pp. 53-61, 2018.

V. SLON ET AL., The genome of the offspring of a Neanderthal mother and a Denisovan father, *Nature*, vol. 561, pp. 113-116, 2018.

D. SCHRIDER ET A. KERN, S/HIC: Robust identification of soft and hard sweeps using machine learning, *Plos Genet.*, vol. 12(3), art. e1005928, 2016.

J. LACHANCE ET AL., Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers, *Cell*, vol. 150, pp. 457-469, 2012.