

Une introduction sur l'utilisation du Deep Learning pour le suivi du comportement animal

Mathieu BONNEAU¹

CORRESPONDANCE

mathieu.bonneau@inrae.fr

Depuis une dizaine d'années, le terme d'intelligence artificielle fait partie du vocabulaire courant, ainsi que, dans une moindre mesure, le terme d'apprentissage profond (deep learning). Depuis bien plus longtemps en revanche, on utilise des modèles calibrés sur des données pour faire des prédictions ou des simulations, afin de mieux comprendre un système ou d'accompagner la prise de décision. Avant l'apprentissage profond, la variable prédite ou simulée devait être explicitement reliée à un ensemble de variables explicatives, à partir desquelles les prédictions étaient faites. Par exemple, on pouvait prédire la quantité de parasites dans une parcelle (variable prédite) en fonction du nombre d'œufs de parasites ayant été déposés (variable explicative 1) et de la météo (variable explicative 2). Construire un modèle s'apparentait alors à (i) choisir les bonnes variables explicatives, c'est-à-dire comprendre biologiquement ou rationnellement de quoi dépend notre variable prédite, (ii) formuler explicitement, via un modèle mathématique, la relation biologique entre ces variables. Aujourd'hui, ce qu'offre l'apprentissage profond, c'est de se passer de cette modélisation explicite du lien entre la variable à prédire et les variables explicatives, mais également de faciliter le choix des variables explicatives.

Prenons pour exemple la prédiction de la posture d'un cochon à partir d'une image. Une approche « pré » apprentissage profond consisterait à détecter les pixels appartenant au cochon sur l'image, puis à calculer des caractéristiques géométriques de la forme créée par ces pixels. L'idée ici est que la posture (variable prédite) dépend de la forme géométrique formée par le cochon (variables

explicatives). Une approche d'apprentissage profond consiste à entraîner un réseau de neurones convolutif pour qu'il puisse directement, à partir de l'image, prédire la posture. Ici, nul besoin de se poser des questions pour savoir quelles sont les informations importantes présentes sur l'image. Le réseau de neurones le fait seul, et dans bien des cas, de manière plus robuste², en perdant néanmoins le pouvoir explicatif des modèles traditionnels.

Mais alors, comment est-ce possible ? Sans entrer dans les détails, les réseaux de neurones peuvent être vus comme des moyens automatiques et intelligents permettant de compresser l'information. Pour le traitement d'images par exemple, le réseau ResNet-50 reçoit en entrée une image de taille 224 x 224 pixels. La couleur de chaque pixel étant codée par des niveaux de rouge, de bleu et de vert, cela représente une information initiale composée de $224 \times 224 \times 3 = 150\,528$ éléments. Après une succession de 141 opérations, ou couches du réseau, cette information est réduite en un vecteur de taille 1 024, soit seulement 0,68 % de l'information initiale. C'est finalement ce vecteur de taille 1 024 qui devient la variable explicative. C'est elle qui est utilisée comme variable d'entrée d'une simple fonction de type logistique, pour prédire la probabilité que le cochon soit dans telle ou telle posture.

Avec les réseaux de neurones, la réduction de l'information se fait de manière automatique, via l'utilisation de fonctions mathématiques simples (les couches du réseau), qui agissent comme des filtres, capables d'extraire ou de porter leur attention sur les zones intéressantes de l'information initiale. Même si ces filtres restent

¹ INRAE - ASSET, UR0143, 97170 Petit-Bourg, Guadeloupe.

² Bonneau, M., Benet, B., Labrune, Y., Bailly, J., Ricard, E., & Canario, L. (2021). Predicting sow postures from video images: Comparison of convolutional neural networks and segmentation combined with support vector machines under various training and testing setups. *Biosystems Engineering*, 212, 19-29.

des éléments mathématiques simples, le type de filtre, leur nombre et leur position dans le réseau restent un sujet de recherche important, dont les résultats les plus utilisés sont bien souvent publiés par des compagnies privées.

Par chance, aujourd'hui, il n'est souvent pas nécessaire de créer de toutes pièces le réseau de neurones permettant de résoudre la problématique liée à son domaine de recherche. Il s'agit plutôt de choisir un type de réseau, puis de l'entraîner sur des données spécifiques.

Il est important de comprendre que les données sont le facteur limitant principal dans de nombreuses applications. Le réseau de neurones est comme un liquide mou, capable de s'adapter à la forme de n'importe quel jeu de données et, virtuellement, de répondre à n'importe quelle requête, pourvu qu'un jeu de données d'entraînement adapté soit disponible. Car derrière les architectures « simples » des réseaux, se cachent des millions de paramètres (25,6 millions pour ResNet-50) constitués par la succession de fonctions simples qui permettent de s'adapter aux données. C'est un point important, car même si l'extraction automatique de l'information semble quelque peu magique, il ne s'agit finalement « que » de s'adapter à un jeu de données adéquat.

Par exemple, nous avons passé plusieurs années à construire un jeu de données permettant de détecter des chèvres créoles sur des images. Le réseau, de type YOLO, fonctionne effectivement très bien grâce aux 6 000 images qui ont été annotées pour l'entraîner. Cependant, dans ce jeu de données, il n'y a aucune chèvre blanche. En conséquence, le réseau présente des performances impressionnantes sur les chèvres de couleur sombre, mais c'est un désastre pour les chèvres de couleur claire. Le défi, dans de nombreux cas, ne porte donc pas sur la conception du réseau,

mais sur la création d'une base de données annotées permettant d'entraîner le réseau à effectuer la tâche demandée. Plus la base de données sera constituée d'exemples provenant de conditions et d'environnements variés, tout en étant annotée avec précision, plus les réseaux créés seront robustes, génériques et utilisables à grande échelle, permettant d'automatiser l'annotation. Dans ce contexte, l'importance des données et de leur partage au sein et entre les différentes communautés ne peut être sous-estimée.

Dans notre domaine, où les environnements sont relativement constants, comme les infrastructures expérimentales, l'intelligence artificielle associée à des capteurs, caméras ou autres, devrait permettre de suivre de manière continue et sur le long terme différents traits de comportement des animaux de ferme, dès lors qu'une quantité suffisante de données annotées sera disponible. Même s'il reste des défis méthodologiques, que nous n'aborderons pas ici, de nouvelles perspectives devraient s'ouvrir, notamment celles des données quantitatives massives, pouvant être qualifiées d'« éthologie quantitative ». Le comportement des animaux deviendra une série de données numériques (séries temporelles) et bénéficiera ainsi d'une nouvelle boîte à outils puissante, celle des modèles mathématiques permettant d'analyser et d'objectiver ce comportement de manière inédite³. Cela s'accompagnera sûrement de son lot de controverses, de savoir-faire qui se créent, qui disparaissent. Mais c'est peut-être finalement la parole que nous donnerons aux animaux d'élevage, en leur permettant de s'exprimer via des séries temporelles, un langage que nous serons capables d'interpréter, sans controverse, qui nous permettra d'aborder d'une nouvelle manière les questions d'émotions ou de bien-être chez les animaux. ■

³ Un exemple, où un réseau de neurones a permis de déterminer ce qui, dans le comportement, permettait de différencier les individus malades des individus sains : Maekawa, T., et al. (2020). Deep learning-assisted comparative analysis of animal trajectories with DeepHL. Nature communications.