

Intelligence artificielle et biologie

Tristan Cazenave

Professeur des universités
LAMSADE

L'intelligence artificielle et en particulier la recherche Monte-Carlo peuvent être utilisées pour concevoir de nouvelles molécules, et notamment des molécules d'ARN. La conception de molécules d'ARN messenger est une des pistes poursuivies pour la conception d'un vaccin, par l'équipe du projet Eterna. Cet article reprend les travaux que nous avons effectués avec Thomas Fournier au LAMSADE dans ce cadre.

La pandémie de la Covid-19 a montré l'importance de pouvoir réagir rapidement face à de nouveaux virus et plus généralement face à de nouvelles maladies. Le but de l'intelligence artificielle est d'automatiser les processus de résolution de problèmes, or la conception de nouveaux traitements pour les différentes maladies fait partie des problèmes qui gagneraient à être automatisés. Pour les applications plus classiques de l'intelligence artificielle comme les jeux ou l'optimisation, les algorithmes ont déjà de meilleures performances que les êtres humains. On pourrait imaginer que des algorithmes de résolution de problèmes accélèrent la découverte de traitements.

Des succès récents de l'intelligence artificielle pour des problèmes combinatoires difficiles comme les jeux de Go, d'Échecs et de Shogi ou encore comme les tournées de véhicules sont en grande partie dus à l'utilisation d'algorithmes de recherche Monte-Carlo. Ce sont des algorithmes qui reposent sur un grand nombre de choix aléatoires servant à explorer l'espace des possibles. Il est tentant d'utiliser ces algorithmes pour résoudre d'autres problèmes combinatoires. Notamment ceux liés à la biologie.



Eterna est un jeu vidéo sur internet inventé par des biologistes. Le but de celui-ci est de trouver des séquences de molécules d'ARN qui ont une forme donnée. Les joueurs choisissent les bases de la séquence de façon à ce que le repliement en deux dimensions de la molécule d'ARN corresponde à la forme prédéfinie. La figure précédente donne un exemple de puzzle d'Eterna. Une base de 100 problèmes similaires issus d'Eterna constitue la base Eterna100 de problèmes de référence pour évaluer les algorithmes de conception de molécules. La conception de molécules ayant une forme donnée a été identifiée par Édouard Bonnet, Pawel Rzazewski et Florian Sikora¹ comme un problème difficile.

En 2018, Fernando Portela, un des contributeurs du projet Eterna a utilisé l'algorithme de recherche Monte-Carlo imbriquée (*Nested Monte Carlo Search*) que j'avais conçu pour les jeux à un joueur afin de résoudre les problèmes d'Eterna100. Complété avec un grand nombre de connaissances

¹ Bonnet, E., Rzazewski, P., Sikora, F.: Designing RNA secondary structures is hard. *Journal of Computational Biology* 27(3) (2020)

biologiques sur l'ARN, cet algorithme a réussi à résoudre automatiquement 95 des 100 problèmes d'Eterna100 alors que les autres algorithmes testés par différents chercheurs en résolvait entre 54 et 72.

Parmi les algorithmes de recherche Monte-Carlo, un des algorithmes le plus efficace est le NRPA (*Nested Rollout Policy Adaptation*). Le principe de NRPA est d'apprendre des probabilités de choisir les actions dans les parties jouées aléatoirement. NRPA a récemment été généralisé vers GNRPA (*Generalized NRPA*), un algorithme de recherche Monte-Carlo capable de prendre en compte des heuristiques représentées comme des probabilités a priori de choisir les actions. GNRPA améliore très nettement NRPA pour les problèmes de tournées de véhicules et pour la conception de molécules. Avec Thomas Fournier nous avons réutilisé dans GNRPA la partie la plus simple des heuristiques utilisées par Fernando Portela dans son programme NEMO (acronyme pour *NEsted MOnTe Carlo RNA puzzle solver*). Nous avons aussi testé de nombreuses améliorations de la recherche Monte-Carlo pour ce problème, et nous sommes arrivés² au final à résoudre 95 des 100 problèmes, le même score que NEMO.

L'algorithme que nous avons utilisé est différent de celui utilisé par Fernando Portela. Il fait appel à beaucoup moins de connaissances du domaine, tout en résolvant le même nombre de problèmes que NEMO, avec des temps de résolution similaires à celui-ci. Une des originalités de notre algorithme est de continuer la recherche à chaque niveau tant que l'algorithme arrive à améliorer sa séquence, une heuristique aussi utilisée dans NEMO et qui est adaptée au problème de conception de molécules d'ARN.

Nous allons continuer ce travail sur la conception de molécules, initié pendant le confinement. Nous travaillons actuellement à généraliser l'approche en appliquant cet algorithme à d'autres problèmes similaires comme la conception de protéines. Nous travaillons aussi à l'amélioration des algorithmes de recherche Monte-Carlo avec comme objectif de résoudre les 100 problèmes d'Eterna100. Le but ultime est d'utiliser ce type d'algorithmes pour créer des molécules utiles pour la découverte de nouveaux traitements. Ainsi, le groupe Eterna travaille actuellement à la conception d'ARN messenger pour lutter contre la Covid-19. Ils ont par exemple lancé en septembre 2020 un défi Kaggle pour trouver automatiquement les instabilités dans les molécules d'ARN messenger. Ceci est important pour la stabilité du vaccin qui est amené à voyager entre son lieu de production et son lieu d'utilisation.

² Cazenave, T., Fournier, T.: Monte Carlo inverse folding. In: Monte Carlo Search at IJCAI (2020)

