

# Intelligence Artificielle et Biologie

Tristan Cazenave

LAMSADE, Université Paris-Dauphine, PSL, CNRS, Paris, France

Tristan.Cazenave@dauphine.psl.eu

La pandémie de 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. 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 [7], d'échecs et de Shogi [8] ou encore comme les tournées de véhicules [5] sont en grande partie dus à l'utilisation d'algorithmes de recherche Monte-Carlo. Il est tentant d'utiliser ces algorithmes pour résoudre d'autres problèmes combinatoires. Notamment ceux liés à la biologie.



Figure 1: Un des problèmes d'Eterna.

Eterna est un jeu vidéo sur internet inventé par des biologistes. Le but 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 1 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, Paweł Rzażewski et Florian Sikora (du LAMSADE) comme un problème difficile [1].

En 2018, Fernando Portela, un des contributeurs du projet Eterna a utilisé mon algorithme de recherche Monte-Carlo imbriquée [2] pour résoudre les problèmes d'Eterna100 [6]. Complété avec un grand nombre de connaissances 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 résolvaient entre 54 et 72.

En mars 2020, au début du confinement, j'ai généralisé un algorithme de recherche Monte-Carlo pour lui faire prendre en compte des heuristiques représentées comme des probabilités a priori de choisir les actions [3]. Pendant ce confinement j'ai cherché comment aider la recherche en biologie. Cela me semblait en rapport avec la lutte contre la Covid-19 et la recherche de traitements même si mes objectifs étaient à plus long terme. J'avais encadré, en 2019, le stage de Thomas Fournier, un étudiant en informatique de Dauphine qui travaillait sur l'amélioration du programme de Fernando Portela. Nous avons alors décidé avec Thomas de travailler ensemble pendant cette période sur la conception de molécules en réponse à la pandémie.

Thomas Fournier a fait une interface entre le programme de Fernando Portela et mes différents algorithmes de recherche. J'ai alors testé de nombreuses améliorations de la recherche Monte-Carlo pour ce problème et nous sommes arrivés au final à résoudre les 95 problèmes que résolvait NEMO (le programme de Fernando Portela, acronyme pour NEsted MOnTe Carlo RNA puzzle solver). L'algorithme que nous avons utilisé est différent de celui utilisé par Fernando Portela et 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 [4]. Une des originalités de cet algorithme que nous avons proposé 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 initié pendant le confinement sur la conception de molécules. Nous travaillons actuellement à généraliser l'approche en appliquant cet algorithme à d'autres problèmes similaires comme la conception de protéines [9]. 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 étant 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 utilisation.

## References

- [1] Bonnet, E., Rzażewski, P., Sikora, F.: Designing RNA secondary structures is hard. *Journal of Computational Biology* **27**(3) (2020)
- [2] Cazenave, T.: Nested Monte-Carlo Search. In: Boutilier, C. (ed.) *IJCAI*. pp. 456–461 (2009)
- [3] Cazenave, T.: Generalized nested rollout policy adaptation. *CoRR* **abs/2003.10024** (2020), <https://arxiv.org/abs/2003.10024>
- [4] Cazenave, T., Fournier, T.: Monte carlo inverse folding. In: *Monte Carlo Search at IJCAI 2020* (2020)
- [5] Cazenave, T., Lucas, J.Y., Kim, H., Triboulet, T.: Monte carlo vehicle routing. In: *ATT at ECAI 2020* (2020)
- [6] Portela, F.: An unexpectedly effective monte carlo technique for the RNA inverse folding problem. *BioRxiv* p. 345587 (2018)
- [7] Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D.: Mastering the game of Go with deep neural networks and tree search. *Nature* **529**(7587), 484–489 (jan 2016)
- [8] Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T.P., Simonyan, K., Hassabis, D.: A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science* **362**(6419), 1140–1144 (2018)
- [9] Vucinic, J., Simoncini, D., Ruffini, M., Barbe, S., Schiex, T.: Positive multistate protein design. *Bioinformatics* **36**(1), 122–130 (2020)