

Optimisation de protocoles d'administration de drogues antitumorales par d'apprentissage *tabula rasa*

Auteur: Éric Pfeiderer

Professeure responsable: Morgan Craig, DMS

20 janvier 2020

1 But

Le but de ce projet est d'adapter et de déployer un algorithme d'apprentissage par renforcement, **AlphaGo Zero**, afin d'optimiser le protocole d'administration de drogues antitumorales. On désire comparer *AlphaGo Zero* avec d'autres méthodes d'optimisation plus traditionnelles, tels que les algorithmes génétiques. Si possible, on désire aussi trouver l'existence de patrons pertinents dans les résultats à l'aide de techniques telles que le regroupement et l'analyse de sensibilité.

2 Description

Les algorithmes d'apprentissages par renforcement sont souvent particulièrement bien adaptés à la prise de décision. On peut modéliser *AlphaGo Zero* comme un agent à la recherche de récompenses dans un environnement dynamique. En interagissant avec son environnement, l'agent tente de développer une politique d'action qui lui permet d'effectuer des choix optimaux afin de maximiser les récompenses obtenues. Dans le contexte des échecs, par exemple, l'agent doit développer une politique qui décrit les meilleurs coups à jouer en fonction de l'état de l'échiquier. Dans le contexte d'essais cliniques virtuelles, l'agent doit plutôt développer une politique d'administration qui maximise l'efficacité du traitement mais minimise son fardeau sur le système immunitaire.

3 Spécifications

AlphaGo Zero emploie un réseau de neurones convolutionnel à deux têtes[1]. La tête principale, dite *policy head*, décide de l'action optimale à prendre en fonction de l'environnement afin de maximiser les récompenses. Le rôle du *policy head* peut facilement être généralisé au contexte d'essais cliniques. La deuxième tête, dite *value head*, sert plutôt à évaluer la qualité de l'environnement et permet de guider l'entraînement du *policy head* par recherche arborescente Monte Carlo (MCTS). La généralisation du *value head* au contexte d'essais cliniques n'est pas aussi simple et demande le développement d'une heuristique cohérente décrivant l'efficacité d'un traitement.

Le modèle de croissance tumorale est fourni par **Cassidy** et **Craig** [2] et on prévoit principalement utiliser le langage de programmation **python**. Notre version d'*AlphaGo Zero* sera développé à l'aide de bibliothèques d'apprentissage automatique non spécialisées telles que **Tensorflow** et/ou **PyTorch**. L'ensemble du projet sera développé et donc accessible sur **Github**[3].

4 Plan de développement

Le projet devrait progresser comme suit:

1. Préparation de l'environnement de travail
2. Implémentation du modèle de croissance tumorale
3. Adapdation et implémentation de la méthode *AlphaGo Zero* au contexte d'essais cliniques
4. Optimisation des hyperparamètres d'*AlphaGo Zero*
5. Entraînement d'*AlphaGo Zero*
6. Analyse de résultats

5

References

- [1] **Silver, D., Schrittwieser, J., Simonyan, K. et al**, Mastering the game of Go without human knowledge, Nature 550, 354–359 (2017) doi:10.1038/nature24270
- [2] **Cassidy T., Craig M.**, Determinants of combination GM-CSF immunotherapy and oncolytic virotherapy success identified through in silico treatment personalization, PLoS Comput Biol 15(11): e1007495. <https://doi.org/10.1371/journal.pcbi.1007495>
- [3] **Pfleiderer E.**, dépôt Github IFT3150, <https://github.com/EricPfleiderer/IFT3150>