



AGUECCI Enzo
CARRON Valentin
7 Avril 2021

Minimisation de regret hypothétique profond et Poker

La théorie des jeux, spécialité de John Nash, réside dans le fait de faire le bon choix. Pour rester dans le domaine des cartes, cette théorie est parfaitement illustrée dans le célèbre film « Las Vegas 21 », où un professeur emmène ses élèves les plus doués en mathématiques à Las Vegas tous les week end, afin de compter les cartes au BlackJack et gagner le plus d'argent possible.

La manière dont le professeur va recruter le personnage principal est la suivante : Kevin Spacey (le professeur) propose à Jim Struggess (l'élève) de choisir entre 3 portes : derrière une seule des trois se trouve un lot.

Jim choisit une porte, Kevin en ouvre une autre et lui montre qu'elle est vide, Jim a donc désormais une chance sur deux de remporter le lot.

Kevin lui propose de rester sur son choix de départ ou changer de porte.

Or, Jim, un brillant mathématicien, sait parfaitement qu'en recevant cette proposition, si il garde sa première idée, il a toujours une chance sur trois de gagner, mais, deux chances sur trois de gagner s'il change d'avis et choisit d'ouvrir la deuxième porte.

C'est ce qu'il fait, et réussit le test de son professeur, c'est ce qu'on appelle le problème de Monty Hall, qui utilise la théorie des jeux.

Dans une première partie nous allons définir plus profondément la théorie des jeux et le poker en général.

Puis nous nous intéresseront à la minimisation de regret (CFR) a travers le Libratus et le Pluribus.

I. LA THEORIE DES JEUX ET LE POKER

1. La théorie des jeux

Pour revenir à l'exemple utilisé dans l'introduction, il peut ne pas être évident qu'il paraisse judicieux, par la peur de se dire « J'aurais du rester sur ma première idée ».

Pour le rendre plus clair, prenons un jeu de 52 cartes classique.

Quelqu'un vous demande de trouver le roi de coeur, vous choisissez la 51ème carte.

La personne retourne les 50 premières et le roi de coeur n'y est pas, c'est donc la 51ème ou la 52ème. Il vous propose de changer de choix et choisir la 52ème si vous voulez, il faudrait être fou de refuser, vous avez en effet $51/52$ soit 98% de chance de trouver le roi de coeur, alors qu'en restant sur votre premier choix vous auriez toujours 1 chance sur 52 !

Les fondements mathématiques de la théorie moderne des jeux sont décrits autour des années 1920 par Ernst Zermelo, et par Emile Borel dans l'article « La théorie des jeux et les équations intégrales à noyaux symétriques ».

Ces idées sont ensuite développées par Oskar Morgenstern et John Von Neumann en 1944 dans leur ouvrage « La théorie des jeux et le comportement économique » qui est considéré comme le fondement de la théorie des jeux moderne.

Le mathématicien John Nash est considéré par beaucoup comme fournissant la première extension significative des travaux de von Neumann et Morgenstern.

Leur but était de réussir à modéliser des jeux où la somme des gains entre les joueurs est toujours nulle, grâce à l'intelligence artificielle.

La théorie des jeux a un large éventail d'applications, y compris la psychologie, la biologie évolutionniste, la guerre, la politique, l'économie et les affaires. Malgré ses nombreuses avancées, la théorie des jeux est encore une science en développement.

Selon la théorie des jeux, les actions et les choix de tous les participants affectent le résultat de chacun.

La théorie des jeux est un cadre théorique pour concevoir des situations entre joueurs concurrents (comme par exemple le poker).

À certains égards, la théorie des jeux est la science de la stratégie, ou du moins la prise de décision optimale d'acteurs indépendants et concurrents dans un cadre stratégique.

Cependant, cette théorie connaît des limites, bien qu'elle crée des modèles d'analyse très performants afin de tirer le meilleur de chaque situation, elle néglige les événements extérieurs, qui ont autant d'influence, si ce n'est plus, que les autres acteurs du « jeu ».

Le Poker

2.

Le poker est un jeu de cartes se jouant à 52 cartes, dans lequel généralement 2 à 8 joueurs s'affrontent autour d'une table.

Le but est d'avoir la meilleure combinaison (cf illust.) en retenant 5 cartes des 7 présentes (2 dans la main + 5 sur la table) ou de réussir à faire coucher les adversaires.

Le Poker, se jouant la plupart du temps dans des casinos, est souvent considéré comme un jeu de hasard, mais il n'en est rien, il s'apparente plutôt à un jeu de stratégie et de prise de décision stratégique, comme par exemple le bluff, soit, donc parfaitement adapté à la théorie des jeux.

Dans ce jeu de carte, à chaque tour de parole une décision est prise et est très importante, plusieurs choix s'offrent au joueur : Fold, Check, Call, ou Raise.

Il est donc difficile de savoir, en fonction de son jeu et des cartes qui peuvent être sur la table quelle option il faut choisir.

De plus si le joueur décide de Raise (relancer), il reste encore à savoir de combien il relance.

Toutes ses décisions ne vont pas être les mêmes en fonction des adversaires, il va donc falloir s'adapter et comprendre comment fonctionnent les personnes autour de la table, et donc à terme, « minimiser les regrets », ce que nous verrons dans la seconde partie.

Décision plus technique qui peut être prise au Poker :

Deux cartes de la même couleur dans les mains, un tirage couleur au Flop, mais la couleur n'est pas touchée à la Turn, ni à la River, beaucoup de jetons sont engagés dans le coup, prendre la décision de tourner la main en bluff est une décision qui pourrait permettre de remporter le pot, en impressionnant l'adversaire.

Classement des mains au poker :									
1.	T ♥	J ♥	Q ♥	K ♥	A ♥				La quinte flush royale
2.	A ♥	2 ♥	3 ♥	4 ♥	5 ♥				La quinte flush
3.	4 ♥	4 ♦	4 ♣	4 ♠					Le carré
4.	3 ♥	3 ♦	3 ♣	+	2 ♣	2 ♥			Le full
5.	Q ♥	4 ♥	9 ♥	2 ♥	6 ♥				La couleur (flush)
6.	A ♥	2 ♦	3 ♣	4 ♣	5 ♥				La suite (quinte)
7.	3 ♥	3 ♦	3 ♣						Le brelan
8.	2 ♥	2 ♦	+	A ♣	A ♠				La double paire
9.	2 ♥	2 ♦							La paire
10.	A ♥								La carte haute (isolée)

II. LA MINIMISATION DE REGRET

Les grandes difficultés pour une intelligence artificielle au poker se comptent au nombre de deux : La dissimulation et la falsification d'informations.

En effet, un joueur n'a pas accès à toutes les informations de ses adversaires et lorsque ceux-ci en donnent, elle peuvent être fausses et c'est ce que l'on appelle du "bluff".

Il faut donc une intelligence artificielle dédiée, dont l'algorithme évolutif permet d'exploiter le peu d'informations dont elle dispose...

La méthode la plus adaptée à ce genre de situation, qui est le plus performant actuellement est donc la minimisation de regret alias "CFR".

La CFR est un algorithme d'entraînement avec soi-même c'est à dire que l'intelligence artificielle apprend en jouant contre elle-même. La «minimisation du regret» signifie qu'à la fin de chaque partie simulée, l'algorithme va revenir en arrière pour explorer toutes les issues possibles. Voir ce qu'il serait arrivé s'il avait joué différemment afin d'en tirer le meilleur profit.

Au départ, la stratégie est aléatoire pour pouvoir ensuite introduire la notion de "regret".

- Le regret peut être soit positif c'est à dire que l'issue de la décision prise n'est pas bénéfique pour l'intelligence artificielle. Il faudra donc prendre une autre décision.
- Le regret peut aussi être négatif et ainsi l'issue de la décision est bénéfique et il faut continuer dans cette direction.

1. Libratus

Libratus est un algorithme qui découle directement de la méthode CFR pour un mode de jeu particulier du poker qui est le heads-up (un contre un).

Il comporte trois principaux modules:

- *La construction d'une stratégie de plan directeur.*

Une solution est pré-calculée pour l'ensemble du jeu en utilisant un programme alternatif.

Cependant pour le heads-up il existe 10^{13} points de décisions unique. L'algorithme utilise alors ce qu'on appelle l'abstraction d'action. En effet beaucoup de ces points de décisions sont semblables et c'est pourquoi on réduit le nombre de décisions possibles afin de créer un jeu plus simple et plus petit tout en conservant les aspects stratégiques du jeu.

Par exemple si un adversaire parie 1,01€ l'intelligence artificielle prendra ce pari comme un pari à 1€ et régira en conséquence.

Une fois l'abstraction construite, un calcul de stratégie de modèle est mit en place pour Libratus en faisant jouer l'intelligence artificielle contre elle-même qui utilise une version améliorée de l'algorithme de minimisation contrefactuelle des regrets de monte Carlo.

La MCCFR va alors enregistrer une valeur de regret pour chaque décision lors de cette expérience et représenter à quelle point il regrette de ne pas avoir choisit telle ou telle action dans le passé.

Alors que de plus en plus de jeux sont simulés, MCCFR garantit que, avec une forte probabilité, le regret moyen d'un joueur pour toute action approche de zéro.

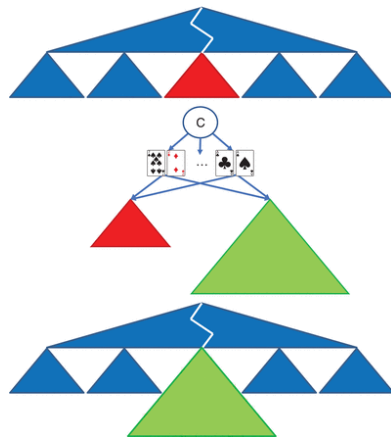
Ainsi, la stratégie moyenne de l'IA sur tous les jeux simulés s'améliore progressivement.

La solution de l'abstraction est ainsi appelé stratégie du plan directeur.

- *La résolution de sous jeux-impliqués.*

Lorsqu'une partie ultérieure du jeu est atteinte pendant le jeu, le deuxième module de Libratus construit une abstraction plus fine pour ce sous-jeu et le résout en temps réel. Contrairement aux techniques de résolution de sous-jeux dans les jeux à information parfaite, Libratus ne résout pas l'abstraction des sous-jeux de manière isolée. Au lieu de cela, il garantit que la solution fine du sous-jeu s'inscrit dans la stratégie de plan plus large de l'ensemble du

jeu. Chaque fois que l'adversaire effectue un mouvement qui n'est pas dans l'abstraction, un sous-jeu est résolu avec cette action incluse.



Un sous-jeu en rouge est atteint pendant la première lecture. Le bleu et le rouge indiquent la stratégie du plan directeur. Le chemin blanc indique la séquence d'action avant le sous-jeu atteint. Une stratégie plus détaillée pour ce sous-jeu est déterminée en résolvant un sous-jeu augmenté dans lequel, à chaque itération, l'adversaire reçoit une main de poker aléatoire et a le choix de prendre la valeur attendue de

l'ancienne abstraction (rouge) ou de jouer dans la nouvelle abstraction plus fine (verte), où la stratégie des deux joueurs peut changer.

Cela oblige Libratus à rendre la stratégie plus fine au moins aussi bonne que celle de l'abstraction originale contre chaque main de poker adverse. En bas du schéma, la nouvelle stratégie est remplacée par l'ancienne.

- *L'amélioration personnelle*

Le troisième module de Libratus est l'auto-amélioration. Il améliore la stratégie du plan directeur en remplissant les branches manquantes dans l'abstraction du plan et calcule une stratégie de théorie des jeux pour celles-ci.

Libratus utilise les mouvements réels de l'adversaire pour suggérer où dans l'arborescence du jeu un tel remplissage en vaut la peine.

Finalement Libratus s'est brillamment imposée face à quatre joueurs professionnels en No-Limit Hold'em, en terminant gagnante de 1,76 million de dollars.

Voici le tableau des pertes des quatre autres joueurs :

Joueur	Perte	Par main
Dong Kim	-84 054 \$	-2,90 \$
Jimmy Chou	-338 347 \$	-11,69 \$
Jason Les	-862 347 \$	-29,80 \$
Daniel McAuley	-275 441 \$	-9,52 \$
Total/Moyenne	-1 560 189 \$	-13,48 \$

2. Pluribus

Pluribus n'est en fait que la version améliorer de Libratus. En effet, Pluribus calcule d'abord une stratégie "BluePrint" en jouant six copies de lui-même, ce qui est suffisant pour le premier tour d'enchères. À partir de là, Pluribus fait une recherche plus détaillée des mouvements possibles dans une abstraction plus fine du jeu. Il anticipe plusieurs mouvements comme il le fait, mais ne nécessite pas de regarder en avant jusqu'à la fin du jeu, ce qui serait prohibitif en termes de calcul. Un nouvel algorithme de recherche à anticipation limitée est la principale percée qui a permis à Pluribus de réaliser un poker multi-joueurs surhumain.

Conclusion

Finale­ment, la méthode de minimisation de regret est une très grande avancée dans le monde de l'intelligence artificielle et de la théorie des jeux. La création de Libratus et de son évolution Pluribus ont vraisemblablement montré que le poker n'est pas seulement un jeu de hasard puisqu'avec des algorithmes tel que la minimisation contre factuelle des regrets de monte Carlo il était possible de ne jamais perdre une partie de poker. Cependant cette évolution peut commencer à poser problème pour les sites de poker en ligne. En effet, des algorithmes se rapprochant de ceux utilisés par Pluribus et Libratus sont de plus en plus utilisés par les joueurs de poker en ligne afin de maximiser leurs gains. Les plateformes de poker se doivent donc de palier ce genre de "tricheries" afin de garder leurs utilisateurs et continuer de fonctionner correctement.