

# Le jeu de poker et l'IA

Jonathan des Rosiers

Université du Québec à Montréal

8 novembre 2017

# Les apprentissages possibles par les jeux

- Apprentissage des actions du jeu
- Planification à court terme
- Planification à long terme
- Coopération
- Résolution de problèmes
- Réflexion stratégique

- Les jeux sont un terrain d'essai (proving ground) pour l'IA
- Les conséquences de nos erreurs sont moins graves

- Une variante du poker
- Origine Texas
- Un des jeux de cartes les plus joués

- Le jeu de poker a des règles simples
- Le poker est un jeu stochastique. Il implique un élément de hasard.
- C'est un jeu à somme nulle
- Le développement d'une ou plusieurs stratégies robustes n'est pas simple, car certaines informations du jeu sont cachées (privées)
- La variante hold'em sans limites à deux joueurs est comparable au jeu de go en terme de noeud de décision :  $10^{160}$  versus  $10^{14}$  avec limite

- Classement pour un ensemble de cinq cartes



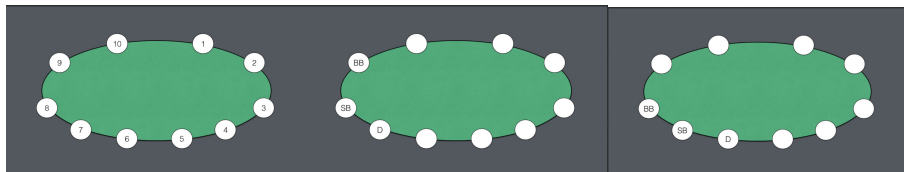


Figure: Position, Small blind et Big blind



hole cards  $\forall$  players

**player actions:**

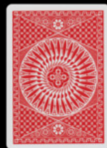
**check, bet,  
raise, fold**



**flop**



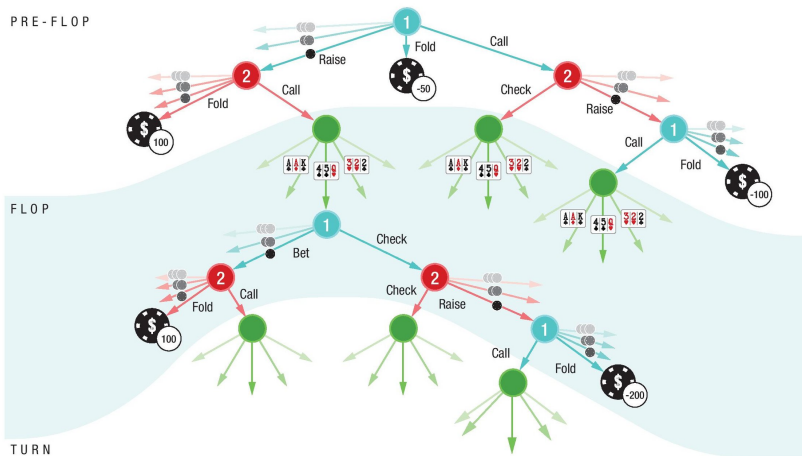
**turn**



**river**



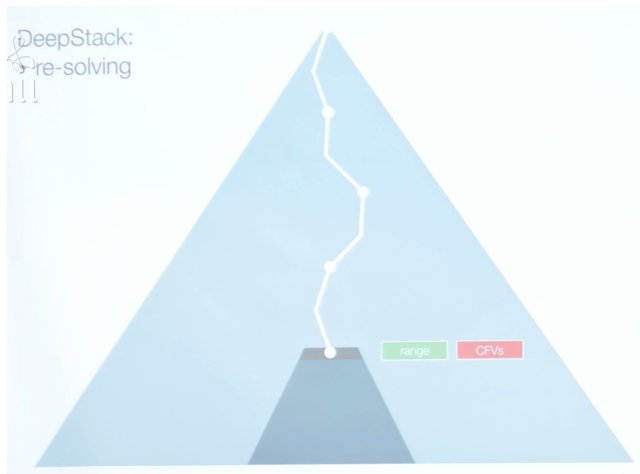
# La représentation



- $\#BB = \text{les jettons en argent}/BB$
- $BB/100 = 1mBB$
- Exemple : Un joueur achète pour 100\$. Le BB est de 10\$. Le joueur a 10BB. Après 1 millions de main, le joueur perd en moyenne 1mBB par 100 mains.

- Gagner contre un groupe de joueurs professionnels
- Résoudre le jeu (le programme ne peut pas perdre)

- Voici les grands principes qui ont guidé les chercheurs pour développer DeepStack

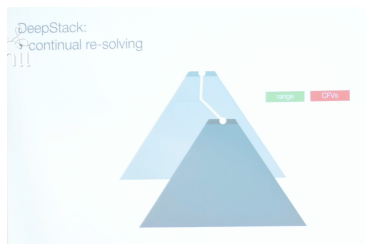


# Pré-résoudre le jeu (DeepStack)

- Joue contre lui-même
- Combinaison de mains jouées (range) (A range consists of a combination of hands)
- Le programme doit maintenir sa combinaison de mains jouées à un état  $x$  du jeu et la valeur contre-factuel(cfv)
- La valeur contre-factuelle correspond à une série de conditions du type "qu'est qui se passerait si?"
  - Plus spécifiquement, la fonction retourne la valeur attendue de l'adversaire atteignant cet état avec une main en particulier
- La valeur cfv et la combinaison de mains jouées sont des vecteurs
- En ayant ces deux vecteurs, il peut oublier tout ce qui est arrivé avant ce point de décision

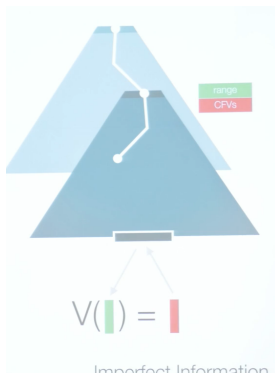
# Résoudre en continu (DeepStack)

- Lorsque le programme fait une action, les deux vecteurs sont mis à jour
- Lors d'un noeud chance (tirage de carte), les valeurs contre-factuelles de l'adversaire sont mises à jour. Il met à jour sa combinaison de cartes(range) en enlevant les cartes du tirage.
- Lorsque c'est à l'adversaire d'agir, aucune de ces deux valeurs n'est modifiée.
- En itérant et en ajustant ces deux valeurs, on arrive à une estimation d'un équilibre de Nash



# La fonction d'évaluation - heuristique (DeepStack)

- DeepStack permet de remplacer ce sous-arbre en utilisant une fonction qui fait une approximation de la valeur contre-factuelle, comme si cet état était résolu avec les cartes de l'itération courante.
- En résolvant les problèmes en continu, il évite d'avoir à maintenir une stratégie pour toute la partie. Il construit la stratégie à chaque fois qu'il doit agir, sans jamais utiliser la stratégie d'un noeud précédent

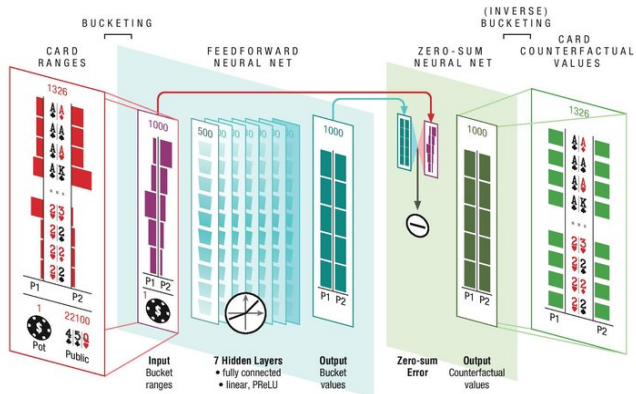


# La fonction d'évaluation - heuristique (DeepStack)

- Les entrées de cette fonction correspondent à la combinaison des cartes(range) jouées des deux joueurs, l'argent dans le pot et les cartes publiques.
- La sortie de la fonction retourne les valeurs de toutes les mains possibles à ce point de la partie.
- 10 millions de mains ont été générées aléatoirement pour entraîner le réseau de neurones créé pour simuler la troisième phase du jeu(le turn).
- La stratégie de DeepStack est statique et est basé sur le résultat des calculs qui ont produit une distribution de probabilités pour les actions possibles



# Réseau de neurones (DeepStack)



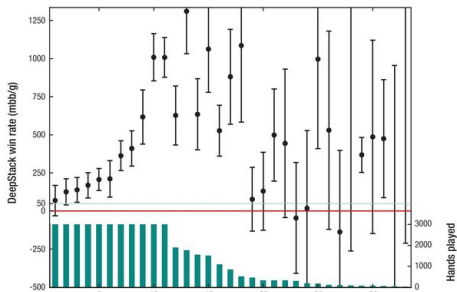
# Exemple(DeepStack)

- Ajustement de la combinaison de mains jouée (range)
- Diminution de la chance
- Ajustement de l'action



# AIVAT(DeepStack)

- Diminution de la variance
- 3 milles mains jouées sont aproximement égales à 90 milles mains jouées sans l'outil AIVAT
- Cette technique a été utilisée pour évaluer les parties jouées contre le groupe de professionnels
- Résultat DeepStack : 494 mbb/g



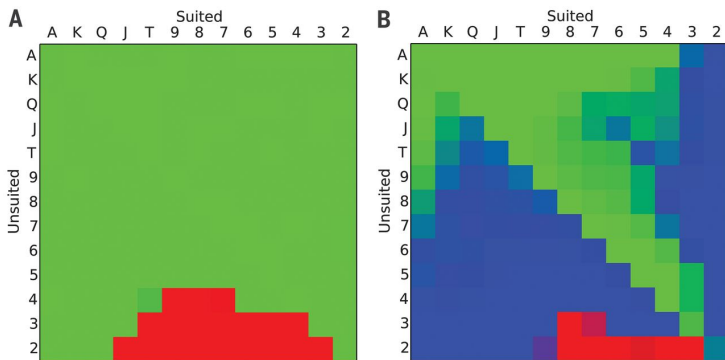
- Une façon de réduire le jeu de poker à un espace plus petit serait de rassembler certaines mains, comme AA et KK et les jouer de la même manière qu'avant le flop (1ere phase du jeu)
- Il y a plusieurs manières de rassembler des mains dans des sous-ensembles

# Regroupement préflop

- Les paires, le regroupement de deux cartes distinctes n'ayant pas la même couleur et le regroupement de deux cartes ayant la même couleur.
- Chaque carte a un rang : 2,3,4,5,6,7,8,9,T,J,Q,K,A
- Chaque carte a une couleur : trèfle,coeur,carreau,pique
- Les cartes (9ca,Kpi) peuvent être représentées (K9cd) et (9pi,Kpi) seraient représentées par K9mc. cd = couleur différente, mc = même couleur
- On a donc 169 mains différentes préflop.

# Regroupement pré flop (Cepheus)

- Comparativement à  $1326((52*51) / (1*2))$  sans le regroupement




- Une meilleure réponse est une stratégie qui obtient la meilleure utilité contre une autre stratégie
- Au poker, cette stratégie correspond à la stratégie qui rapporte le plus d'argent contre tous les joueurs dont la stratégie a été révélée et doit rester statique. En un contre un, la meilleure réponse correspond à la meilleure stratégie contre le joueur  $x$ .
- Plus un agent est "exploitable", plus la meilleure réponse va gagner de l'argent contre celui-ci.

- L'équilibre de Nash est un profil de stratégies basé sur le fait qu'aucun joueur ne peut augmenter son utilité (rendement) en changeant sa stratégie.
- DeepStack permet de faire une approximation d'un équilibre de Nash.
- Deepstack charge sa stratégie pour le noeud courant. Même si cette stratégie est chargée dynamiquement, elle demeure statique, mais stochastique.



# Recherche limitée en profondeur dans l'arbre de décision résolu (DeepStack)

- Deepstack doit faire une recherche en profondeur, avec une limite sur la profondeur, pour éviter d'avoir à raisonner jusqu'à la fin du jeu
- Avec une limite sur la profondeur de 4 action, on diminue les points de décision de  $10^{160}$  à  $10^7$ .
- Action : fold, bet, 2bet ou 3bet et all in.
- Prise de décision en : <5s en utilisant une Geforce GTX 1080

 Matej Moravcik, Martin Schmid, Neil Burch, Viliam Lisy, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling(2017)  
DeepStack: Expert-Level Artificial Intelligence in No-Limit Poker

 Site web consulté le : 1 nov. 2017  
<http://poker.srv.ualberta.ca/preflop> (range Cepheus)  
<https://speakerdeck.com/tsantero/deepstack-expert-level-artificial-intelligence-in-heads-up-no-limit-poker>

<https://www.deepstack.ai/>

Fin