

Résumé de l'article Probabilities and Simulations in Poker

February 2015

1 Introduction

Loki est un jeu de poker développé et mis à jour à l'Université d'Alberta depuis 1997. L'intelligence artificielle de Loki a pour but de gagner le plus d'argent qu'il peut dans n'importe quelle situation.

1.1 Loki-1

Loki-1 est le premier niveau de Loki. L'intelligence artificielle de cette version a le niveau d'un joueur intermédiaire.

Le but de la thèse était d'améliorer la stratégie de Loki-1 et sa façon de profiler les joueurs.

1.2 Loki-2

Loki-2 : Première amélioration : Une meilleure stratégie pour déterminer quelle action est la plus intéressante. Le problème étant que cette stratégie est facile à prédire. Ce calcul retourne trois probabilités pour chaque action (se coucher, enchérir/checker, raise/parier). L'IA choisit de façon aléatoire l'action à effectuer. Après cette amélioration, Loki-2 a écrasé Loki-1.

2 Poker

Dans ce chapitre, les règles du jeu de poker sont expliquées et les travaux d'autres chercheurs sur le sujet sont présentés.

3 Loki-1

3.1 Architecture

La méthode de Loki-1 pour profiler perd de l'efficacité si l'adversaire montre des signes d'agressivité pendant le jeu. Cette méthode est meilleure si tous les

adversaires ont seulement checké ou enchéri. l'Opponent Modeler (OM) modifie le poids de la table d'un adversaire si tous les adversaires ont seulement checké ou enchéri. l'OM modifie le poids de la table d'un adversaire après qu'il ait observé une de ses actions en fonction du contexte du jeu. Maj les probabilités pour toutes les mains est un procédé appelé "reweighting". Après l'action de chaque adversaire, l'OM appelle "l'évaluateur de main" une fois pour chaque main possible et modifie le poids dans chaque cas pour qu'il soit en accord avec les nouvelles informations.

3.2 Betting strategy

3.2.1 Preflop expert system

Quand c'est la première fois que Loki joue dans le preflop, Loki utilise un système basé sur les règles pour sélectionner une des quatre stratégies de preflop définies. Il détermine le nombre de paris que Loki va enchérir et sous quelles conditions cela sera parié/raise. Cette sélection est basée sur la moyenne des retours sur investissement de toutes les cartes de Loki. Il calcul quelle stratégie sera la meilleure (de la plus agressive à la plus passive) en fonction de toutes les paramètres de la partie.

3.2.2 Postflop

Pour avoir la qualité d'une main après le flop, Loki combine la force et le potentiel de la main dans une variable appelée "effective hand strength" (EHS). C'est une estimation qui donne la probabilité que la main courante soit actuellement la plus forte ou qu'elle le devienne quand la carte suivante est découverte.

$$EHS = \text{hand_strength} + (1 - \text{hand_strength}) * \text{hand_potential}$$

(Cf un exemple dans le paragraphe en question du mémoire).

Le potentiel d'une main peut être positif ou négatif. Positif (PPOT) : la probabilité qu'une main devienne la meilleure.

$$PPOT = (\text{be_behind_end_ahead} + \text{be_behind_end_tied} / 2 + \text{be_tied_end_ahead} / 2) / (\text{total_be_behind} + \text{total_be_tied} / 2)$$

Négatif (NOPT) : probabilité qu'une main soit mauvaise

$$NPOT = (\text{be_ahead_end_behind} + \text{be_ahead_end_tied} / 2 + \text{be_tied_end_behind} / 2) / (\text{total_be_ahead} + \text{total_be_tied} / 2)$$

Les deux sont calculés avec toutes les mains adverses possibles et les possible cartes qui seront découvertes au prochain tour.

la stratégie basique du postflop est basée sur deux paramètres : un seuil de "postflop-raise" et un seuil de "postflop-call". Si l'EHS est meilleur ou égal au postflop-raise alors, il raisera quand moins de deux paris auront été faits pendant le round et enchérira sinon. Quand l'EHS est meilleur ou égal au postflop-call, il pariera si pesonne d'autre n'a parié et enchérira sinon, sauf quand ils sont deux ou plus parieurs à enchérir et quand il n'a pas déjà enchéri un pari à ce round.

Quand l'EHS est inférieur au postflop-call ou que sa décision est de suivre, il y a l'option du semi-bluff, appelé avec "pot odds", et appelé avec "showdown odds" est aussi considérée.

Semi-bluff : parier si personne ne l'a encore fait dans le round actuel et qu'on a un bon PPOT pour pouvoir parier et enchérir à la fois.

Pot odds : ratio du montant d'argent dans le pot par rapport à la quantité nécessaire pour enchérir le pari actuel. Quand Loki utilise pot odds, il va garder sa main si ses chances de gagner sont meilleures que les retours attendus du pot.

Showdown odds : ratio du montant d'argent attendu par le showdown pour être dans le pot.

3.3 Opponent modeling

L'OM assigne un tableau de poids à chaque adversaire, indexé par les deux cartes du la main de départ. Tant que Loki connaît ses deux cartes et les trois cartes de flop, il y a 1081 entrées dans le tableau après le flop. Chaque fois qu'un adversaire fait une action de pari, le poids pour cet adversaire est modifié pour tenir compte de l'action effectuée et des cartes en commun. L'OM garde des statistiques pour chaque adversaire entre les parties. Ces données sont utilisées pour calculer la fréquence de suivi, enchère et raising pour chaque adversaire par round et le nombre de paris effectués pour enchérir. De cette fréquence, l'OM déduit la moyenne et la variabilité des seuils attendus pour une action en question. (calcul ratio : Cf page 32).

4 Probability triples (Loki-2, profilage dynamique)

Pour améliorer Loki-2, ils ont choisi d'utiliser "probability triples" (PT), un ensemble composé de trois probabilités représentant les trois types d'actions, pour avoir une stratégie randomisée. Un PT est une liste ordonnée de trois valeurs. $PT=(f,c,r)$ (fold, call, raise) avec $f+c+r=1.0$. Chaque valeur représentant la probabilité que l'action suivante dans un état donné soit "se coucher", "suivre" ou "relancer". Cette technique est utilisée dans trois états de Loki-2: 1. en tant que stratégie autonome 2. en tant que facteur de repondération dans le module de modélisation de l'adversaire 3. en tant que mécanisme de génération d'action durant la simulation.

4.1 Probability triple generation function

La fonction de génération PT décrit comment un player devrait se comporter avec une certaine paire de cartes dans une situation spécifique.

$$\begin{aligned} f &= P(\text{action=fold} \mid \text{pair of cards and game context}) \\ c &= P(\text{action=call} \mid \text{pair of cards and game context}) \\ r &= P(\text{action=raise} \mid \text{pair of cards and game context}) \end{aligned}$$

Les fonctions utilisent l'évaluateur de main avec une stratégie basée sur des règles définies par des experts pour calculer les trois valeurs. L'évaluation d'une

main comprend la force (la probabilité que la main soit la plus forte) et le potentiel (la probabilité que la main devienne la plus forte après que les cartes communes soient retournées) d'une main.

La première version de PT était plus simple que la stratégie de Loki-1 mais a montré l'avantage d'avoir une stratégie non-déterministe. Le plus gros avantage d'une stratégie non-déterministe étant que Loki-2 peut choisir de façon aléatoire ses actions, en se basant sur la liste de probabilités plutôt qu'en suivant la simple action retournée par la stratégie de Loki-1.

Pour générer le PT pour une main, la valeur de la main est d'abord calculée. C'est une estimation de la probabilité de gagner. Cette valeur est utilisée par un enregistrement de règles pour calculer les probabilités de se coucher, suivre ou de relancer.

$S()$: donne les informations publiques d'une partie

h : une main

$EHS(h, S())$: la valeur de la main

$PT_{s,ehs}(f, c, r)$: représente le PT

$$PT_{s,ehs}(f, c, r) = \begin{cases} (0, .25, .75) & \text{si } EHS(h, S()) > .75 \\ (.20, .80, 0) & \text{si } .75 \geq EHS(h, S()) > .50 \\ (.50, .40, 0) & \text{otherwise} \end{cases} \quad (1)$$

Cf page 38, l'algorithme pour une fonction de génération PT simplifiée, utilisant 4 règles.

4.2 Using probability triples

4.2.1 As a betting strategy

A chaque fois que Loki-2 avait à jouer, il appelait le PT générateur de fonction et sélectionnait son action en se basant sur le PT retourné. Le choix était ensuite fait en générant un nombre aléatoire compris entre 0.0 et 0.1. Un seul nombre aléatoire est généré au début de chaque main et est utilisé à chaque fois qu'il faut choisir une action. Ce nombre aléatoire est gardé car un définit le niveau de mise ou de relance quand la partie avance. De ce fait, le jeu de Loki-2 est plus agressif. Son agressivité devrait être consistante tout au long d'une partie car ce n'est pas une bonne idée ce ne est pas une bonne idée de miser fortement au début du jeu. Utiliser un simple nombre aléatoire permet de fixer le style de jeu de Loki-2 dans une partie. L'utilisation du générateur de fonction PT permet au jeu de Loki-2 d'être plus imprévisible.

4.2.2 As a reweighting factor

Le module de modélisation de l'adversaire contient un tableau pour chaque adversaire contenant les poids pour chaque main qu'il peut avoir. ce poids est la probabilité que le comportement passé d'un adversaire donne une paire spécifique de cartes.

$wt = P(\text{observed actions} \mid \text{pair of cards})$

Les triplets de probabilités sont utilisés pendant le postflop comme un facteur de repondération pour mettre à jour la table de poids de chaque adversaire. Loki-2 calcule le PT pour chaque main que l'adversaire peut avoir et multiplie le poids de chaque main par la valeur du triplet qui correspond à l'action que l'adversaire a effectuée. Les valeurs du triplet permettent à Loki-2 de faire face aux informations douteuses données par l'adversaire.

4.3 Experiments

4.3.1 Design

Dans cette partie, les résultats qui ont été obtenus en faisant jouer Loki-1 contre Loki-2 sont expliqués. Le style de jeu d'un joueur est défini par le pourcentage de mains jouées (relâché / conservateur) et la fréquence de relance (agressivité / passivité).

4.3.2 Results

Cette partie contient les résultats obtenus

5 Selective sampling simulation

5.1 How simulation works

Quand c'est au tour de Loki-2 de jouer, il utilise la simulation habituelle pour avoir une estimation du pourcentage de check ou de sur-enchère. Se coucher est considéré comme avoir 0% de chances parce qu'il n'y a pas d'autre résultat. Chaque expérimentation est jouée au moins deux fois pour considérer la conséquence d'un check ou d'un call et une fois pour observer la conséquence d'un bet/raise. Dans chaque cas, le taux d'argent gagné ou perdu est déterminé et moyenné avec les résultats correspondants à l'expérience. A la fin de la simulation, les moyennes des deux enregistrements d'expériences sont comparés.

Il y a ensuite dans le texte une explication de la simulation "parfaite", en choisissant les cartes du jeu à l'avance etc...

5.2 Dealing cards out

Dans cette partie, ils expliquent qu'ils ont choisi de pouvoir choisir au début de la partie les mains des adversaires et les cartes du milieu pour faire les simulations.

5.3 Experiments

In this section they explain the results they had, with the choose of the cards.

5.4 Comments about selective sampling simulation

5.4.1 advantages

The simulation-based approach used in Loki-2 has experimentally proved to be better than Loki-1's static approach. Selective sampling simulations discover information that improves the values obtained by the static evaluation function.

5.4.2 Simulation trade-offs

Dans cette partie, il est expliqué comment la simulation pourrait être meilleure.

5.4.3 Comparison with alpha-beta

Alpha-beta est un algorithme efficace pour le profilage de deux joueurs. Cet algorithme est comparé dans cette partie à l'algorithme "selective sampling simulation".

5.4.4 Summary

Dans cette partie il y a un résumé du chapitre.

6 Other examples of selective sampling simulation

Dans ce chapitre, les algorithmes de simulation utilisés dans les réseaux de simulation et les jeux sont passés en revue.

7 Conclusion and future work

Utiliser les triplets de probabilités dans Loki-2 lui a permis d'avoir une meilleure stratégie et cela a présenté une amélioration significative dans le jeu de Loki-2, face aux anciennes versions de Loki et face à des êtres humains. Face à des être humains, les meilleurs résultats étaient obtenus quand les trois améliorations effectuées étaient utilisées. Loki-2 gagne plus d'argent que les anciennes versions de Loki. Cependant, le programme ne joue pas encore au niveau des championnat du monde.