

Amélioration du stratège par Apprentissage



Apprendre la fonction d'évaluation

- Définition de la fonction d'évaluation

$$\text{Eval}(s) = w_1 f_1(s) + w_2 f_2(s) + \dots + w_n f_n(s)$$

- Apprendre les w

Apprendre la fonction d'évaluation

- Définition de la fonction d'évaluation

$$\text{Eval}(s) = w_1 f_1(s) + w_2 f_2(s) + \dots + w_n f_n(s)$$

- Fonction dépendante ou non du temps (du n° de coup)

$$\text{Eval}(t, s) = w_1(t) f_1(s) + w_2(t) f_2(s) + \dots + w_n(t) f_n(s)$$

- Apprendre les w ou les $w(t)$

Apprendre à explorer ?

- Paramétrisation de l'algorithme de recherche
 - Algorithme à profondeur fixée
 - Tendance à favoriser des prises importantes au dernier coup
 - Algorithmes à profondeur variable
 - Nécessite un critère d'arrêt
 - Stabilité de la situation
 - Validité de la fonction d'évaluation
 - Algorithmes dépendants des contraintes temporelles
 - Prise en compte du temps
 - Profondeur fixée dépendante du temps

Comment apprendre ?



Mise au point du stratège

- A la main
- Par apprentissage supervisé
- Par apprentissage faiblement supervisé

Apprentissage supervisé

- Principe

- Apprentissage à partir d'une base d'exemples
 - On fournit un ensemble de situations étiquetées par un expert
 - Apprentissage automatique de la fonction d'évaluation $f(s)$
 - Algorithmes « simples »
 - Type réseaux de neurones

Apprentissage faiblement supervisé

- Principe : on apprend à partir de parties entières gagnées/perdues
 - On fait jouer la machine contre elle-même des (centaines de) milliers de parties
 - On raffine au fur et à mesure le stratège
 - But : trouver des poids optimaux
 - Problème : Credit assignment problem
 - Déterminer les coups responsables de victoire/défaite
 - Plus généralement les bons et les mauvais coups
 - Comment changer les poids ?

Apprentissage supervisé

Apprentissage à partir
d'exemples (etat,coup)



● Méthodes

- Apprentissage à partir d'un ensemble de paires $E = \{(etat_i, coup_i), i=1..N\}$
 - correspondant à des coups « idéaux » dans un ensemble de situations

Apprentissage à partir d'un ensemble d'apprentissage

- Base de coups idéaux $E = \{(\text{etat}_i, \text{coup}_i), i=1..N\}$
 - On cherche la fonction f qui va choisir dans chaque situation etat_i le coup coup_i
 - Il faut que l'ensemble d'apprentissage soit représentatif de toutes les situations
- Généralement algorithmes itératifs

- Pb : Avoir l'ensemble E

- +sieurs possibilités :

1. Parser des parties diffusées sur Internet
2. Stocker la courbe $f(s)$ sur une partie et déterminer automatiquement un bon coup
 - Par exemple celui à partir duquel $f(s)$ a été toujours positif

Apprentissage faiblement supervisé

Apprentissage sur des parties entières



1. Apprentissage par exploration



Exploration simple de l'espace des fonctions d'évaluation

- On cherche un point $W=(w_1, w_2, \dots, w_n)$ optimal dans l'espace des W .
- Idée
 - Explorer +- exhaustivement l'espace des (w_1, w_2, \dots, w_n)
 - On part de W^0
 - A une itération t donnée, on a un vecteur W^t
 - On veut modifier le vecteur pour une composante j
 - On fait jouer un joueur utilisant W^t contre
 - un joueur utilisant $W^t + (0, \dots, \varepsilon, \dots, 0)$ où $\varepsilon > 0$ apparait sur la j -ieme composante
 - un joueur utilisant $W^t - (0, \dots, \varepsilon, \dots, 0)$
 - On sélectionne le meilleur,
 - On recommence pour le même j jusqu'à ce qu'on n'améliore plus
 - On recommence pour un autre j

Algorithme

- Initialiser W aléatoirement (ou d'après vos essais)
- Tant que ...
 - Pour $j=1$ à p (nombre de sous fonctions)
 - Évaluer $(W | w_j + \varepsilon)$ et $(W | w_j - \varepsilon)$. Et déterminer le sens de l'amélioration
 - Tant qu'il y a amélioration faire
 - $w_j = w_j + \varepsilon$ ou $(w_j = w_j - \varepsilon)$
- Evaluation : Par rapport à un joueur de référence