Amélioration du stratège par Apprentissage

Apprendre la fonction d'évaluation

• Définition de la fonction d'évaluation Eval(s) = $w_1 f_1(s) + w_2 f_2(s) + ... + w_n$ $f_n(s)$

Apprendre les w

Apprendre la fonction d'évaluation

Définition de la fonction d'évaluation

Eval(s) =
$$w_1 f_1(s) + w_2 f_2(s) + ... + w_n$$

 $f_n(s)$

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

Eval(t, s) =
$$W_1(t) f_1(s) + W_2(t) f_2(s) + ... + W_n(t) f_n(s)$$

Apprendre les w ou les w(t)

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= {(etat_i, 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
 - Plusieurs possibilités :
 - 1)Parser des parties diffusées sur Internet
 - 2)Disposer d'un Oracle qui nous indique les coups à préférer

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 $W=(w_1, w_2, ..., w_n)$ optimal dans l'espace des W.
- Idée
 - Explorer + ou exhaustivement l'espace des (w₁, w₂, ...,w_n)
 - On part de W⁰
 - A chaque itération t :
 - On fait une modification légère de W^t et on évalue l'efficacité du nouveau joueur par rapport à l'ancien.
 - Si le nouveau joueur gagne, non garde le changement. Sinon on revient à l'ancien vecteur de paramètres.

Evaluation de l'efficacité d'un déplacement

- Plusieurs possibilités
 - X parties par rapport à un joueur de référence
 - Celui qui a le meilleur ratio de gain l'emporte
 - X parties faisant s'affronter le nouveau joueur avec l'ancien
 - + Evolutif
 - Risque de divergence
 - Mix entre ces deux possibilités
 - Adaptation du joueur de référence tous les K déplacements