

狸王 アピール文書

ザイオソフト コンピュータ将棋サークル

野田久順 岡部淳

鈴木崇啓 河野明男

目次

- 狸王
- 主な工夫点
 - NNUE評価関数
 - 強化学習における教師局面生成
 - 思考ノード数の固定
 - 棋譜生成器の探索パラメータの調整
 - 探索パラメーターの自動調整
 - クラスタリング
- 使用ライブラリ

狸王

野田「『〇〇〇〇〇〇〇〇〇〇タヌポン』っていう
参加名にしようと思うんですけど…」

岡部「…」

那須「…第6回電王トーナメントに使う
予定だった名前のほうが良いと思います」

野田「…ああ、…そうでしょうか」

狸王(続)

- もし第6回電王トーナメントが開催されていたとしたら、この名前で出場する予定でした
- ディフェンディングチャンピオン(=王者)という意味が込められています

NNUE評価関数

- 狸王は評価関数にNNUE評価関数を使用しています
 - ネットワーク構造に
halfkp_256x2-32-32を使用しています

強化学習における教師局面生成

- 狸王では評価関数を強化学習により強化しています
 - 「教師データが不足した環境での機械学習結果改善手法」と「elmo式学習法」を用いています
- 教師局面の生成に工夫を加えています
 - 思考ノード数の固定
 - 棋譜生成器の探索パラメータの調整

思考ノード数の固定

- これまでの教師局面生成では、探索深さを固定して自己対局を行い、棋譜を生成してきました
 - 探索深さを深くした場合、対局の終盤において、思考ノード数が極端に増え、思考時間が長くなる場合があります
- 狸王では思考ノード数を固定し、思考時間が長くなりすぎるのを抑制しています
 - 棋譜生成時間が予測しやすくなりました

棋譜生成器の探索パラメータの調整

- 棋譜生成器の探索パラメータを、自己対局のパラメータに合わせて最適化しています
 - 後述の手法でパラメータの自動調整を行っています
- これにより、より質の高い教師局面が得られるようになったと考えています

探索パラメーターの自動調整

- WCSC26で導入した、
探索パラメータの自動調整を行っています
 - hyperoptで探索パラメータと
自己対局の勝率のサンプリングを行い、
Gaussian Processを用いて
なるべく良さそうなパラメータを選んでいきます

探索パラメーターの自動調整（続）

※WCSC27PR文書より抜粋・改変

- 探索パラメーターを入力、ベースとなる思考エンジンに対する勝率を出力とする関数を考えます
- 関数の値を最大化する探索パラメーターを探します
 - 大域的最適化問題にモデル化できます
- 自己対戦により得られる勝率には大きなノイズが含まれています

探索パラメーターの自動調整（続）

- Tree-structured Parzen Estimator (TPE)
 - 機械学習のハイパーパラメーターの最適化に用いられるアルゴリズムの1つです
 - PythonライブラリHyperoptに実装されています
 - 1探索パラメーターセットあたり48対局、数百パラメーターセット分の自己対戦の勝率をサンプリングします
 - Hyperoptの出力をそのまま使ったらダメ！
 - Hyperoptは最高の値を出力した最初の値を出力します
 - サンプルの誤差が大きい場合、本来弱いはずの探索パラメーターが最終結果として出力されてしまう場合があります

探索パラメーターの自動調整（続）

- Gaussian Process (GP)
 - ノンパラメトリックな回帰分析などに用いられています
 - sklearn等の実装されています
 - Hyperoptで得られたサンプルをGaussian Processでノンパラメトリック回帰分析します
 - 探索パラメーターに対する勝率の関数の確率分布が得られます
 - サンプルの中で関数の確率分布の平均値が最も高いものを最終的な解として出力します

探索パラメーターの自動調整(続)

- 前回からの変更点1
 - サンプルング初期の、ランダムサーチによるサンプルング回数を増やしています
 - これにより、局所最適解に陥りにくくなると考えています

探索パラメーターの自動調整(続)

- 前回からの変更点2
 - 自己対戦を思考ノード数を固定して行っています
 - 時間を固定した場合、各思考エンジンに割り当てられるCPUリソースがばらつき、勝率が50%に寄る問題があります
 - 思考ノード数を固定することにより、この問題を解決しました
 - 合わせて、同時対局数を物理CPUコア数から論理CPUコア数に増やし、サンプリング精度を上げています

クラスタリング

- WCSC27でPonanzaが導入した「eXtream Lazy Smp」を改良したものを実験中です
- 他のクラスタリング手法と比較し、レーティングが高かったものを採用予定です

使用ライブラリ

- Apery
 - tanuki-wcsc28開発時点で評価関数が最も強かったため
評価関数をランダムスタートで機械学習させるための
教師局面の生成に使用しました
- やねうら王
 - 独自の工夫を加えるにあたり、改造しやすく、
レーティングも高いため、使用しました
- tanuki-
 - 強化学習のもとになる評価関数として使用しました

よろしくお願いします