みざうら王 with お多福ラボ PR 文書

みざうら王チーム

特長

・dlshogi 互換エンジンを作った(dlshogi の model ファイルがそのまま使える)
　本家 dlshogi より優れている点もいくつかある。

・詰将棋ルーチンとして df-pn を改良したものを実装した。メモリと時間が無限にあるなら、どんな長編でも間違わずに解ける。(実際にはメモリと時間は無限にないので実際は 50 手を超えるものは現実的には解けないこともあるが…)

・ソースコードは GitHub で公開しているのでそれを見て欲しい。
　　https://github.com/yaneurao/YaneuraOu

学習手法

学習はいわゆる強化学習なのだが、どのような教師をどういう配分で混ぜれば良いのかは知られていない。
・序盤と中終盤の局面の割合は？
・戦型はどれくらいバラけているのが好ましい？
・序盤はどれくらいバラけているのが好ましい？
・入玉将棋の割合はどれくらいが好ましい？
・どれくらいの po(playout)で対戦させた棋譜が好ましい？

そこで、いま公開されている AobaZero、dslhogi の教師局面、それから floodgate の R3600 以上の棋譜、水匠の入玉絡みの棋譜、自己対局で生成した棋譜(800po,1600po,2400po,…)など複数の教師データを用意して、それぞれを 1 本の stream とみなし、1epoch 分の学習に用いる局面数は 500 万と固定化して学習させることにした。

つまり、AobaZero : dlshogi : 水匠 ＝4:5:1 のように書けば、その配分から成る 500 万局面分の一時ファイルが生成される。(無論、前回用いた続きの箇所から)　この一時ファイルから学習する。(ようにした)

epoch が進んだときに、最適な配分は変化していくものと思われる。

しかし、それを試すにも 10 epoch ぐらいは学習させて、対局させるなり loss(損失関数の値)や accuracy(検証用データとの指し手一致率)などを比較しないといけない。そのような比較をしていると手戻りするので時間がもったいない。

かと言ってそのような検証なしだと、悪い配分のまま学習させてしまうことになるので、本来その DNN の architecture の持つポテンシャルまで至らない。

そこで我々が考えたのは、**天国と地獄メソッド**である。

PC は複数台所有しているので、条件(教師の配分)をそれぞれ変えて同時に学習を進める。

そのなかで一番成績の良かったものを生かして(**天国**)、残りの PC の学習結果は捨てる(**地獄**)。

一番成績の良かった PC のデータを残りの PC にもコピーする。そしてまた条件をそれぞれ変えて同時に学習を進める。

それを繰り返す。

こうすることにより、手戻りなしで学習ができる…はずである。

これを書いているのは 3 月 23 日。正直、当日までに学習が終わるかどうかわからない。時間が足りなさそうなら、GCP で A100 を借りて学習を回す予定である。

// この PR 文書は選手権後に差し替える(かも知れない)