Selene　ＰＲ文書

開発者 西海枝 昌彦（さいかいし　まさひこ）

【なにで学習しているの？】

方策勾配法というアルゴリズムにより機械学習を行っています。

【方策勾配法ってなに？】

方策勾配法とは強化学習に分類され、自己対局した後、勝ったプレイヤーの指し手に報酬を与えて、次からは指しやすくし、負けたプレイヤーの指し手は逆に指しにくくするように評価値を調整していきます。

【なにが良いの？】

自己対局する場合には、ある程度のランダム性が求められます。ランダム性が無い場合、コンピューターはいつも全力で指してしまうため、そのときの評価値が「居飛車有利！」という値になっていた場合、居飛車しか指さなくなります。全部の棋譜が居飛車だけだと、振り飛車の戦型を学習することができません。

どのように“ある程度のランダム性”を実現するか、大変悩むところなのですが、方策勾配法の場合はアルゴリズムとしてランダムに指すことが組み込まれていて、気にせず大丈夫！

また、自己対局する際のプレイヤーの強さを選ぶことができ、最強プレイヤー（そのときの最善手を非常に高い確率で指す）から弱いプレイヤー（あからさまにダメな手は指さないが、次善手、３番手、・・も指す可能性がある）まで、設定が可能です。

【どのように学習しているの？】

プロ棋士の棋譜約５万局を「自己対局した棋譜」として、学習させた後、自己対局によって強化を行います。

自己対局時には強いプレイヤーと弱いプレイヤーを混ぜて行い、強い同士対局することもあれば弱い同士で対局することもあります。

「プロ棋士の棋譜５万局のみ学習」させたものよりも、「さらに１０万局の自己対局を行ったもの」のほうが勝率 65%程度になります。

今回のプログラムでは、大会開始日まで自己対局の回数を増やしていき、挑もうと思っています。よろしくお願いします。