

GP によるゲームの評価関数生成のための事例獲得

—ゲームの対戦に基づく事例の信頼度の学習—

堀之内 剛史

河口 信夫

稲垣 康善

(名古屋大学工学部)

1 はじめに

思考ゲームのプログラムにおいて、指し手をゲーム木探索により選択する場合、局面の評価関数を適切に定めることが重要である。筆者らは、思考ゲーム-棋を対象として、評価関数を学習により獲得することを目標とし、進化的プログラミング (GP) を用いて、評価関数を生成する枠組を提案した [1]。この枠組では、相対的に優勢のついた局面の値が事例としての判定されることを前提としている。しかし、複雑なゲームを対象とした場合、局面の優勢を生成するものが容易でない。本研究では、ゲームの対戦により事例を獲得する手法を提案する。GP により生成した二つの異なる評価関数間での対戦結果は、事例自体の良し悪しを反映していると考えられる。事例に正しく優劣がついていると信頼できる度合いとして、信頼度という概念を導入し、ゲームの対戦結果から、事例の信頼度を学習する。

2 諸定義

定義 2.1 事例を二つのゲームの局面の組 (ψ, ϕ) で表す。先手にとって、局面 ψ は局面 ϕ よりも有利である場合、この事例を正事例、そうでない場合、負事例と呼ぶ。

定義 2.2 評価関数 f によって、事例 (ψ, ϕ) の二つの局面を評価した場合に、 $f(\psi) > f(\phi)$ の関係が成り立つとき、評価関数 f は事例 (ψ, ϕ) を満たすという。

定義 2.3 事例の集合 S を事例ベースとする ($|S| = n$)。評価関数 f が満たす事例の集合を $I_f(S)$ と表す。評価関数 f が満たさない事例の集合を $\bar{I}_f(S) = S - I_f(S)$ と表す。

定義 2.4 事例 $(\psi, \phi) \in S, i \in \{1, 2, \dots, n\}$ に対して信頼度 $credit(i)$ を割り当てる ($0 \leq credit(i) \leq 1$)。信頼度の初期値は 0.5 とする。

3 信頼度の学習

強いゲームプレイプログラムであるほど、正しく局面を評価している。そこで、二つの異なる評価関数を用いたプログラム間で対戦した場合、勝ち方の評価関数は負け方の評価関数よりも局面を正しく評価していると考えられる。すなわち、GP により生成した二つの異なる評価関数においては、勝ち方の評価関数が満たす事例は正事例である可能性が高く、満たさない事例は負事例である可能性が低い。事例の信頼度の学習は、以下の 3 つの操作を学習の 1 サイクルとし、その繰り返しによって行われる。

操作 1 事例ベース S に含まれる各事例の信頼度に基づいて、二つの評価関数 $f, g (f \neq g)$ を GP により生成する。

操作 2 それらの評価関数を用いたプログラム間で対戦する。先手勝、後手勝を交代して二試合を行う。

操作 3 対戦結果にしたがって、各事例の信頼度を更新する。

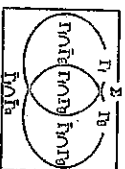


図1: 事例ベースの表現

【操作 3: 信頼度の更新】

定義 2.3 から、事例ベース S は $I_f, I_{\bar{f}}, I_g, I_{\bar{g}}, I_f \cap I_g, \bar{I}_f \cap I_g, I_f \cap \bar{I}_g, \bar{I}_f \cap \bar{I}_g$ の四つの集合に分割される (図 1)。操作 2 の結果にしたがって、各集合ごとで事例の信頼度を更新する。

勝ち方の評価関数が満たす事例は、信頼度を上げる。

$$(例) \text{credit}(i) = \text{credit}(i) + \lambda_1 (1 - \text{credit}(i)) \quad (0 < \lambda_1 < 1) \quad \dots (1)$$

負け方の評価関数が満たさない事例は、信頼度を下げる。

$$(例) \text{credit}(i) = \text{credit}(i) - \lambda_2 \text{credit}(i) \quad (0 < \lambda_2 < 1) \quad \dots (2)$$

$$I_f \cap I_g, I_f \cap \bar{I}_g, \bar{I}_f \cap I_g \text{ に含まれる事例は、信頼度を何に大きく更新する。}$$

(例) (1), (2) それぞれの式で、 $\lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 < 1$ を満たす λ_3, λ_4 を用いる。

4 実験と評価

対象ゲームは tic-tac-toe (三目並べ) とした。このゲームのすべての可能な局面をゲーム木探索により系統的に評価して、正事例 150 例、負事例 150 例をランダムに生成した。この合計 300 例を事例ベースとして与え、各事例の信頼度を 200 サイクルにわたり学習させた。学習過程における正事例、負事例それぞれの信頼度の平均値の推移を図 2 に示す。学習終了後、正事例と負事例それぞれについて、信頼度 0.1 刻みでの度数分布を図 3 に示す。

図 2 では、負事例の信頼度の平均値は初期値 (0.5) からそれほど減少していないが、これは、信頼度が大きな値をとる負事例が多いためである (図 3)。仮に信頼度が 0.1 以下の事例を事例ベースから排除すると、正事例は 150 例のうち 142 例、負事例は 72 例が残り、学習前は負事例が事例ベース全体の 50% を占めるのに比べて、学習後は 34% 弱に減少する。正事例、負事例を正類に識別するには至っていないが、任意に与えた事例ベースから正事例を識別するのに、本手法が有用であることが確認できた。

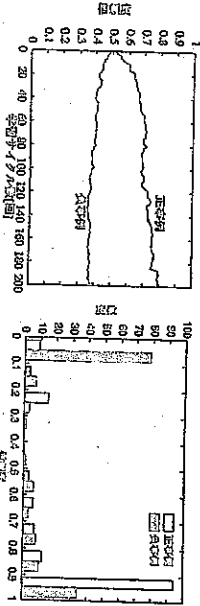


図2: 正事例と負事例の信頼度の平均値の推移

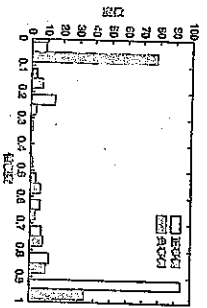


図3: 学習後の正事例と負事例の信頼度の度数分布

参考文献

- [1] 堀之内, 河口, 稲垣: 進化的プログラミングを用いたゲームの局面評価関数の生成, 第10回人工知能学会全国大会, pp.649-652, 1998.