Eric Siegel, Columbia University

Astro Teller, Carnegie-Mellon University

Edgar Learns to Play Othello

Edgar, an intelligent computer program, is one of two leading fictional characters (the other is a human) in the novel, "Exegesis", by Astro Teller. "Exegesis" is a very entertaining and insightful novel, and truly is the first book I have read in one sitting since the decline of my attention span at age 13.

Edgar is also a non-fictional, automatically-learned Othello player developed by the same Astro Teller. Currently, you can play head to head with Edgar (follow the link, "BEAT Edgar"), where he sits patiently as part of an advertisement for "Exegesis." Note that if you follow this link, you'll see, "But watch out--the more you play, the more Edgar learns!" This is NOT TRUE -- Edgar was trained off-line (i.e., in the past), as described on this web page, and is not learning any more.

Teller devised a clever learning method that is specifically designed for Othello. It takes a unique approach to the temporal credit assignment problem of machine learning. Temporal credit assignment is how a system deals with delayed reinforcement, where you don't know how bad or good each move was, only what the final outcome was. Teller's approach contributes by representing temporal credit explicitly, and without imposing a manually-designed method to judges intermediate gameboard states. Usually, implicit measurements to handle temporal credit assignment are the norm, e.g., the approach to Checkers in chapter one of Tom Mitchell's "Machine Learning" text.

The experiments by Teller that lead to Edgar's strategy are preliminary, and we only describe the approach here to a limited degree of detail. However, Edgar plays Othello fairly well, beating most but not all people, and presenting a challenge for other automatic methods to beat. For details on comparisons to other methods and to a few human players, see the results of Project Othello, for which 18 student projects used genetic programming to develop automatic Othello players. Note that there is also an intro to computer science assignment that uses EDGAR as a benchmark.

Teller's system played 70,000 games to learn a set of weights for a three dimensional matrix. The matrix assigns a weight for each board position (two dimensions) and each move number, i.e., time (the third dimension -- bet you thought it was the fourth dimension). The weights are used to determine each move as follows:

For each choice Edgar has to make while playing a game, of all possible legal moves, it picks the one with the maximum value for the following little formula:

- The learned time-space value (row, column, time) PLUS
- Number pieces that would be flipped for that move TIMES
- Some pre-assigned weight specific to that time-space location (cf. last paragraph, below)

There is one exception to this strategy. The program does 1 step look ahead ONLY TO SEE IF MOVING THERE MAKES THE CORNER A LEGAL MOVE FOR THE OPPONENT. If so, then a learned value is added to that move's score and the decision making proceeds as normal.

Each of the 70,000 games during which those weights (Edgar's "brain") were being learned were played against one of the following simple, fixed strategies:

- Edgar with noise
- "Random play" strategy
- "Always flip as many pieces as possible" strategy

- Edgar style strategy, but ignoring the "flip-count" and with the following fixed values for (row, column, time) with time ignored:

```
64 -8  8  8  8  8 -8 64
-8 -8  0  0  0  0 -8 -8
 8  0  0  0  0  0  0  8
 8  0  0  0  0  0  0  8
 8  0  0  0  0  0  0  8
 8  0  0  0  0  0  0  8
-8 -8  0  0  0  0 -8 -8
64 -8  8  8  8  8 -8 64
```

During learning, at the end of each game, the results were trickled back through the time dimension of the brain matrix. Read the previous sentence out loud to your spouse. What this means is that depending on how bad Edgar lost a game, all the moves it made that game were penalized at their respective times (i.e., move numbers) in the matrix. Or, if Edgar won, positive the moves are rewarded (i.e., the value is increased).

Regarding the third component of the equation above, Astro said, "I meant to do some hill-climbing over those weights, but never got around to it, so the 1s and 0s you see are just the default values that I never learned better values for. I have no doubt that Edgar would play at least a little better if some simple hill-climbing (over further random games) was done on those weights."

email: *evs at cs dot columbia dot edu*