Composing Efficient, Robust Tests for Policy Selection

Dustin Morrill, Thomas J. Walsh, Daniel Hernandez, Peter R. Wurman, Peter Stone

A real AI deployment problem: Hundreds of candidate deployment policies, dozens of test cases, but you can only run a few test cases...

Test case $t \in T$ measures a particular skill

Test cases $\langle t_i, s_i \rangle$ collection of test cases $t$ and weights $s$

Test score $\mathbf{A}$ weighted average of the test case results.

Robust Test Construction

Target policies $T$ and case $\#1 \#2 \#3 \#4 \#5 \ldots$

$C_1$ 0 1 1 1 1 1

$C_2$ -1 0 1 1 1 1

$C_3$ -1 -1 0 1 1 1

$C_4$ -1 -1 -1 0 1 1

$C_5$ -1 -1 -1 -1 0 1

Test cases (opponents) 16 and 41 were chosen by CVaR $(h)$: RPOSST SEQ: 41 provides a nearly 50/50 information split and 16 (a weaker policy) is beaten soundly by only very good policies.

The 3 best policies identified with just these 2 tests are the bluest columns (strongest policies) in the 46 X 46 matrix.

Empirical Results

Gran Turismo 7 1v1

500 policies, 96% holdout, $m=3$

Gran Turismo 7 1v1 experiment

Possible Test max error size

Uniform 0 5

Strongest (C5) 1.4 1

Middle (C3) 0.6 1

[0.0, 0.5, 0.0, 0.5] 0.2 2

[0.27, 0.0, 0.47, 0.27] 0.07 3

RPOSST Selections