Revisiting *Safe Strategies for Agent Modelling in Games*

Michael Bowling
Department of Computing Science
University of Alberta
bowling@cs.ualberta.ca

An Improved Proof of Theorem 1

The following theorem appears in McCracken and Bowling [2004].

Theorem 1. *As $T \to \infty$, the worst case average reward of following the Safe Policy Selection algorithm will be at least that of the safety policy.*

The original proof relied on a fact that is not at all self-evident: the worst-case is achieved when the opponent maximally and myopically exploits $\pi^{(t)}$ at every time step. While this is true, showing it amounts to proving Theorem 1 itself. The proof below is a more direct, non-circular proof.

Proof. First, we need to show that $\epsilon^{(t)} > 0$ for $t = 1, 2, \ldots$. This is evident by induction on t. When $t = 1$, $\epsilon^{(t)} = f(1) = \beta > 0$. Assume it holds for t. Then, we know $\pi^{(t)}$ is $\epsilon^{(t)}$-safe, so for all $a_{-i}^{(t)}$, $V(\pi^{(t)}, a_{-i}^{(t)}) - r^* \geq -\epsilon^{(t)}$. Therefore, $\epsilon^{(t+1)} \geq f(t+1) = \frac{\beta}{T+1} > 0$.

Looking at the definition for how $\epsilon^{(T)}$ is computed, we can recursively apply the definition to get,

$$\begin{aligned} \epsilon^{(T)} &= \epsilon^{(T-1)} + f(T) + V(\pi^{(T-1)}, a_{-i}^{(T-1)}) - r^* && (1)\\ &= \sum_{t=1}^{T} f(t) + \sum_{t=1}^{T-1} (V(\pi^{(t)}, a_{-i}^{(t)}) - r^*) && (2) \end{aligned}$$

We know that $\pi^{(T)}$ is $\epsilon^{(T)}$-safe, so

$$\begin{aligned} r^* - V(\pi^{(T)}, a_{-i}^{(T)}) &\leq \epsilon^{(T)} && (3)\\ &= \sum_{t=1}^{T} f(t) + \sum_{t=1}^{T-1} (V(\pi^{(t)}, a_{-i}^{(t)}) - r^*) && (4) \end{aligned}$$

By rearranging and collecting the sums,

$$\begin{aligned} \sum_{t=1}^{T} (V(\pi^{(t)}, a_{-}^{(t)}i) - r^*) &\geq -\sum_{t=1}^{T} f(t) && (5)\\ \frac{1}{T} \sum_{t=1}^{T} V(\pi^{(t)}, a_{-}^{(t)}i) &\geq r^* - \frac{1}{T} \sum_{t=1}^{T} f(t) && (6) \end{aligned}$$

In the limit as $T \to \infty$, the right-hand-side approaches r^*, and thus the left-hand-side is at least the safety value. □

Acknowledgements

Thanks to Sam Ganzfried for noting the circular argument in the original proof.

References

Peter McCracken and Michael Bowling. Safe strategies for agent modelling in games. In *AAAI Fall Symposium on Artificial Multi-agent Learning*, October 2004.