Discovery, Utilization, and Analysis of Credible Threats for 2 X 2 Incomplete Information Games in TOM

(Extended Abstract)

Jolie Olsen
University of Tulsa
Tulsa, OK
jolie.d.olsen@gmail.com

Sandip Sen
University of Tulsa
Tulsa, OK
sandip@utulsa.edu

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

General Terms

Design, Management, Theory

Keywords

Theory of Moves, Threats, Games of incomplete information

1. TOM BACKGROUND

Steven Brams's Theory of Moves (TOM) is an alternative to traditional game theoretic treatments of real-life interaction in which players choose strategies based on analysis of future moves and counter-moves that arise if game-play commences at a specified start state and either player can choose to move first [1]. Unlike classical normal form games, TOM requires the initial state of a game to be determined as a function of initial strategies selected by players.

1.1 Rationality Rules

Upon establishing an initial state, the following rules dictate play in TOM:

• Either player can unilaterally switch its strategy, and thereby change the initial state into a new state, in the same row or column as the initial state.

• In response to the player moving first, the other player can unilaterally switch its strategy, thereby moving the game to a new state.

• Alternating responses continue until the player whose turn it is to move next chooses not to switch its strategy. The game terminates in a final state, deemed the *outcome*.

The decision to switch or not from a current strategy is made using backwards induction. Application of the above rationality rules result in convergence to *Non-Myopic Equilibria* (NME).

1.2 Threat Power

In TOM, the power asymmetries of players can be analyzed using *threat power*. Assume a player p_1 wields power against its opponent p_2 and wishes to induce an outcome (i, j) that would not normally occur in normal game play. p_1 can threaten its opponent. A threat is *compellent* if the threat and breakdown strategy are the same and *deterrent* if they are different where *breakdown state* is the Pareto inferior outcome it is threatening and *threat state* is the Pareto dominated outcome it is trying to induce by using a threat.

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

2. PROTOCOL FOR THREAT POWER IN INCOMPLETE INFORMATION GAMES

2.1 Overview

In a complete information setting, where players have full knowledge of the payoff structure of the game, credible threats can be easily identified. If an agent lacks knowledge of its opponent's payoffs, identifying credible threats requires additional interactions and careful analysis. Knowledge of one's own payoff structure and game NME is not always sufficient to deduce the full payoff matrix. However, an equivalence relation on the 2×2 games exists which preserves credible threats, and hence an agent need only identify the equivalence class of the current game and thereafter can analyze any representative game in the class to identify credible threats in the current game. To identify the equivalence class, an agent needs its payoff values and the computation effort, "e-values", which we define as the number of unilateral strategy switches necessary for equilibria convergence. The following protocol, partitioned into a learning phase and an inference phase, provides an agent this knowledge.

2.2 Learning Phase

Ghosh and Sen proposed a learning algorithm in which players with incomplete knowledge sets, through repeated play, converge to a single outcome [2]. While most often these learners converge to an NME, the algorithm provides no insight for games which are cyclic and hence e-values cannot be obtained. An enhanced version of the algorithm to procure e-values is presented with the following procedure:

• Each agent selects an initial strategy, thus determining the subgame played

• Each agent calculates $P_j^{s,p}(S_i)$, the probability of player p_j moving from state S_i in subgame with initial state/player combination (s,p). For a player p_k, two distinct P values can be calculated: (1) The probability that p_k's *opponent*, denoted $p_{\bar{k}}$, will move from S_i; calculated as the ratio of times $p_{\bar{k}}$ has moved from S_i rather than stayed, initialized to 1.0 to provide for initial exploration, and (2) the probability

p_k itself will move, calculated with the following formula:

$$P_k^{s,p}(S_i) = \sum_{l=i+1}^{4} \left(\prod_{n=i+1}^{l-1} P_k^{s,p}(S_n) \right) (1 - P_k^{s,p}(S_l)) f(O_l^k, O_i^k)$$

where O_l^k and O_i^k are p_k's payoff at outcomes O_l and O_i, respectively; and f is a function constructed to reflect p_k's preference between two outcomes.

• Play terminates when a cycle is reached or both players choose not to move.

2.3 Inference Phase

Given knowledge of payoff values and e-values, an agent can identify the equivalence class of the 2 × 2 ordinal value game (the four outcomes ranked 1 through 4 by each player) game it is playing as follows: (1) it generates a set of *belief variables* for each state of the game representing its opponent's preference for that state, (2) the following *prediction rules* dependent upon e-values are used to eliminate values from these sets and establish inequalities, where $e(\alpha)$ are the e-values for subgames with initial state α; $e(\alpha_k)$ the distinct e-value for subgame with initial state/player combination (α, p_k); $P^k(\alpha)$ is p_k's payoff at α; α^D is the outcome diagonal to α, α_k^{ND} (p_k's "direct neighbor") the result of a unilateral strategy switch by p_k, and α_k^{NI} (p_k's "indirect neighbor") the result of a unilateral strategy switch by $p_{\bar{k}}$.

1. $e(\alpha) = <4,4> \iff \alpha$ is the mutually most preferred.
2. $e(\alpha_k) \in \{0, 4\} \iff \alpha$ is not p_k's least preferred state.
3. $e(\alpha_k) \in \{1,2,3\} \Rightarrow \alpha$ is not p_k's most preferred state.
4. If α, β, γ are distinct outcomes and $e(\alpha_k) = e(\beta_k) = e(\gamma_k) = 0 \Rightarrow$ the remaining state is p_k's least preferred state.
5. If $e(\alpha_k) = 0$ and $e(\alpha_k) = 4$ and for all remaining states β, $e(\beta_k) \in \{1,2,3\}$, $\Rightarrow \alpha$ is p_k's most preferred state.
6. $e(\alpha_k) = 4$ or $e(\alpha_k^{NI}) = 3 \Rightarrow \alpha$ is one of p_k's two most preferred states.
7. If $e(\alpha_k) = 0$ and $P^k(\alpha) = 2 \Rightarrow$ the game does not contain a mutually least preferred state.
8. If α is $p_{\bar{k}}$'s most preferred state, and $P^k(\alpha^D) < P^k(\alpha) \Rightarrow e(\alpha) = <4,0>$
9. If $P^k(\alpha^D) < P^k(\alpha)$ and $e(\alpha) = <4,0> \Rightarrow \alpha$ is one of $p_{\bar{k}}$'s two most preferred states. Furthermore, if α is $p_{\bar{k}}$'s second most preferred state $\Rightarrow \alpha^D$ is $p_{\bar{k}}$'s most preferred state.
10. If $P^k(\alpha^D) < P^k(\alpha)$ and $e(\alpha_k) = <4,2> \Rightarrow \alpha$ is $p_{\bar{k}}$'s most preferred state.

(3) if Step 2 does not yield the equivalence class, an agent uses e-values to generate P-values and derive inequalities used in conjunction with those collected in Step 2 to solve a CSP to reduce belief variables until the equivalence class has been found.

3. ANALYSIS OF THREAT POWER

While on surface level all threats might appear the same, scrutiny reveals certain threats might be more effective than others. We propose the following categories for classifying threats: Pure Superfluous (threat state is the unique NME), Learning Induced Superfluous (threat state is the Pareto dominating NME out of two NMEs), Detrimental (threat state is Pareto inferior to an existing NME), or Useful (none of the above). As their names suggest, superfluous threats suppose a player will fare no better inducing threat power as opposed to adhering to basic TOM rules, and detrimental threats should have a negative impact on utility.

Empirical results from simulations justify the preceding classification scheme. Two sets of simulations were performed: ones where threat power was used and ones where it was disregarded. Threat power runs consisted of two phases: *learning/inference phase* (threatener employs the mechanism detailed to identify existing credible threats), and *threat phase* (threatener utilizes said threats). For runs disabling threat power, simulations consisted of the same number of iterations, but with both agents depending solely on Enhanced TOM Learners to form strategies. Three initial strategy selection methods were used. In random strategy selection (R), agents choose initial strategy at random. In exploratory Learning (EL), p_k gradually learns the behavior of its opponent and subsequently selects its initial strategy by calculating for all available strategies s_k^i the probability of selecting s_k^i via the following equation, then choosing the one with the highest probability:

$$P(s_k^i, p_k, t) = \begin{cases} \frac{1}{1+e^{-2t/I_p}} & : s_k^i = \arg\max u(s, p_k) \\ \frac{1}{|S|} & : s_k^i \neq \arg\max u(s, p_k) \end{cases}$$

where I_p is the number of iterations remaining in the current phase and $u(s, p_k)$ calculates the utility of p_k when selecting initial strategy s.[1] Finally, a hybrid method employs (R) during learning/inference and (EL) in the threat phase.

As predicted, superfluous threats showed no added improvements for agents using threat power as opposed to EL; in fact, while threat power usage increased utility over R, EL produced the same (and better) results. Detrimental threat usage resulted in utility decrease for the threatener, with a surprising result being that the greatest utility increase during simulation, nearly 38%, was incurred by the threatenee when detrimental threats were used against it. Overall, useful threats worked well for threateners, increasing utility upwards of 19%. A marked difference between EL versus threat power is in the impact on social welfare for games with useful threats. While EL increases utility of *both* agents, often by commensurate amounts for the threatener, threat power induces a negative impact on the threatenee, resulting in a 25% utility decrease in some games.

4. REFERENCES

[1] S. J. Brams. *Theory of Moves*. Cambridge University Press, 1994.

[2] S. Sen and A. Ghosh. Theory of moves learners: Towards non-myopic equilibrium. *AAMAS*, 2005.

[1] $P(s,p,t)$ allows for exploration as the initial strategy is selected randomly on the first iteration and subsequent strategies are selected by highest utility with a monotonic increasing probability - a sigmoid function dependent upon how many iterations agents are set to play.