Discovering theorems in game theory: Two-person games with unique Pure Nash Equilibrium payoffs

by Pingzhong Tang and Fangzhen Lin

Tomasz Klochowicz

Advanced Topics in Computational Social Choice 2021
Institute for Logic, Language and Computation
University of Amsterdam

18.10.2021

Overview

Project's idea

The paper is a part of a bigger project on discovering theorems in game theory and computational social choice using computers.

It is focused on using computers to discover new classes of two-person games that have **unique Pure Nash Equilibria payoffs**.

P. Tang, F. Lin, Computer aided proofs of arrow's and other impossibility theorems, Artificial Intelligence 173 (2009) 1041–1053.
P. Tang, F. Lin, Two equivalence results for two-person strict games, Games and Economic Behavior (2011) 479–486.
F. Lin, P. Tang, Computer aided proofs of arrow's and other impossibility theorems, in: AAAI'08, 2008.

Methods – overview

1. **Formulate** the notions from game theory in First Order Logic. A class of games corresponds to a first order sentence. ✍
2. **Prove** that universal sentences are sufficient conditions for all games to satisfy certain property iff they are sufficient for all 2×2 games. (analogically to the preservation theorem) ✍
3. **Generate** all sentences of an interesting form. 💻
4. **Check** if any of them is sufficient condition for 2×2 game to satisfy the property. 💻
5. **Collect** the weakest conditions. 💻
6. **Make sense** of the conditions. Which classes of games they correspond to? What theorems were proved? ✍

Two person games in strategic form

Let $\mathbf{N} = \{1, 2\}$ be the set of **players**.

Let A and B be the **sets of actions** of the players. We use a, b, a', b' etc. to denote a single action of a player.

Let $\leq_1$ and $\leq_2$ be the total orders on the set of profiles: $A \times B$. We call them **preference relations** of the players. Let's use $<_i$ and $\simeq_i$ intuitively.

Observe that preference relations correspond to the utility function in normal form: $(a, b) \leq_1 (a', b')$ iff $u_1((a, b)) \leq u_1((a', b'))$

For each action $b \in B$ we define $A^\star(b)$ to be the **set of best responses** to action b by player A:

$$A^\star(b) = \{a \in A \wedge \forall_{a' \in A}(a', b) \leq_1 (a, b)\}$$

Two person games in strategic form

Two profiles (a, b) and (a', b') are said to be **payoff equivalent** iff for all $i \in N$:

$$(a', b') \simeq_i (a, b)$$

A profile (a, b) is a **Pure Nash Equilibrium** of a game iff:

$$a \in A^\star(b) \land b \in B^\star(a)$$

A game has a **unique PNE payoff** iff all the PNEs of that game are payoff equivalent.

Interesting classes of games

A game is **strictly competitive** if players preferences are weakly opposite i. e. if for every pair of profiles (a, b) and (a', b'):

$$(a, b) \leq_1 (a', b') \text{ iff } (a', b') \leq_2 (a, b)$$

A game is **strict** if for both players, different profiles have different payoffs i.e. for all $i \in N$:

$$(a, b) = (a', b') \text{ iff } (a, b) \simeq_i (a', b')$$

Two person games in FOL

To define two person games in FOL we need to ensure that the **preference relations are total orders**. Let Σ denote the set of the following sentences for each player $i \in N$:

(1) **Reflexivity**: $\forall_{a,b}(a, b) \leq_i (a, b)$

(2) **Strong connexivity**: $\forall_{a,b,a',b'}(a, b) \leq_i (a', b') \vee (a', b') \leq_i (a, b)$

(3) **Transitivity**: $\forall_{a,b\ldots}(a, b) \leq_i (a', b') \wedge (a', b') \leq_i (a'', b'') \Rightarrow (a, b) \leq_i (a'', b'')$

Observe that all these conditions can be easily rewritten in CNF for 2×2 games.

Two-person finite games correspond to the first-order finite models of Σ.

Classes of games in FOL

We can define a class of games by adding more conditions to Σ:

(4) **Strict:** $\forall_{a,b\ldots}(a,b) \simeq_i (a',b') \Rightarrow (a = a' \wedge b = b')$

(5) **Unique PNE payoff:** $\forall_{a,b\ldots} NE(a,b) \wedge NE(a',b') \Rightarrow (a,b) \simeq_1 (a',b') \wedge (a,b) \simeq_2 (a',b')$

(6) **Strictly competitive:** $\forall_{a,b\ldots}(a,b) \leq_1 (a',b') \equiv (a',b') \leq_2 (a,b)$

Where:

$NE(a,b)$ iff $\forall_{a,b}\left[\forall_{x\in A}((x,b) \leq_1 (a,b) \wedge \forall_{y\in B}((a,y) \leq_2 (a,b))\right]$

We know that $\Sigma \models (6) \Rightarrow (5)$ (Osborn and Rubinstein 1994).
Let's generate more theorems of this kind.

M.J. Osborne, A. Rubinstein, A Course in Game Theory, MIT Press, (1994).

The preservation theorem

Theorem 1

To prove that any two player game satisfying universal condition Q, has unique PNE payoff it suffices to prove that any 2×2 game satisfying that condition has unique PNE.

Intuition behind the proof:

1. Suppose that a "big" game does not have a unique PNE payoff.
2. There are (at least) two profiles (a, b) and (a', b') in that game such that they violate uniqueness.
3. Consider the 2×2 game $A = \{a, a'\}$ and $B = \{b, b'\}$. It still satisfies Q and does not have a unique PNE payoff. ✔

Framework - formulas

The paper is focused on the conditions similar to the strictly competitive game's condition:
$(a, b) \leq_1 (a', b') \equiv (a', b') \leq_2 (a, b)$.

Since $p \equiv q$ can be expressed in CNF as $(\neg p \vee q) \wedge (p \vee \neg q)$ Then, by taking expressions of the form $(a, b) \leq_i (a', b')$ to be literals, we consider the propositions of the following form:

$$([(a_1, b_1) \leq_1 (a_2, b_2) \vee (a_3, b_3) \leq_1 (a_4, b_4)] \wedge [(a_5, b_5) \leq_1 (a_6, b_6) \vee (a_7, b_7) \leq_1 (a_8, b_8)])$$

Moreover, we allow for negations in front of each literal which leads to 810 000 conditions. The list can be pruned further using logical dependencies (entailment/subsumption) We check them against 1950 2×2 games.

Raw results

When program finds a condition that satisfies uniqueness it does not check stronger conditions. Therefore it only returns the weakest conditions for uniqueness. It has found seven conditions:

$$(x_1, y) \leqslant_1 (x_2, y) \supset (x_2, y) \leqslant_2 (x_1, y) \wedge (x, y_1) \leqslant_2 (x, y_2) \supset (x, y_2) \leqslant_1 (x, y_1),$$
$$(x_1, y) \leqslant_1 (x_2, y) \supset (x_1, y) \leqslant_2 (x_2, y) \wedge (x, y_1) \leqslant_2 (x, y_2) \supset (x, y_2) \leqslant_1 (x, y_1),$$
$$(x_1, y) \leqslant_1 (x_2, y) \supset (x_2, y) \leqslant_2 (x_1, y) \wedge (x, y_1) \leqslant_2 (x, y_2) \supset (x, y_1) \leqslant_1 (x, y_2),$$
$$(x_1, y_1) \leqslant_1 (x_2, y_1) \supset (x_1, y_2) \leqslant_2 (x_2, y_2) \wedge (x, y_1) \leqslant_2 (x, y_2) \supset (x, y_1) \leqslant_1 (x, y_2),$$
$$(x_1, y) \leqslant_1 (x_2, y) \supset (x_1, y) \leqslant_2 (x_2, y) \wedge (x_1, y_1) \leqslant_2 (x_1, y_2) \supset (x_2, y_1) \leqslant_1 (x_2, y_2),$$
$$(x_1, y_1) \leqslant_1 (x_2, y_2) \supset (x_1, y_1) \leqslant_2 (x_2, y_1) \wedge (x_1, y_1) \leqslant_2 (x_2, y_2) \supset (x_2, y_1) \leqslant_1 (x_2, y_2),$$
$$(x_1, y_1) \leqslant_1 (x_2, y_2) \supset (x_1, y_2) \leqslant_2 (x_2, y_2) \wedge (x_1, y_1) \leqslant_2 (x_2, y_2) \supset (x_1, y_1) \leqslant_1 (x_1, y_2).$$

Observe that the strictly competitive games condition is not in the results.

Analysing the results - Example

$$(9)\ (a_1, b) \leq_1 (a_2, b) \Rightarrow (a_2, b) \leq_2 (a_1, b)$$
$$\wedge$$
$$(a, b_1) \leq_2 (a, b_2) \Rightarrow (a, b_2) \leq_1 (a, b_1)$$

"looks like condition for strictly competitive games, except that the strategy of one of the players is fixed in each implication."

A change of an action that (weakly) increases the payoff of the acting player (weakly) decreases the payoff of the other player.

That clause corresponds exactly to the class of *weakly unilaterally competitive games* as indicated by Kats and Thisse (1992).

If a game is strictly competitive then it is also weakly unilaterally competitive

A. Kats, J.-F. Thisse, Unilaterally competitive games, International Journal of Game Theory 21 (3) (1992) 291–299.

Analysing the results - Example

$$(10)\ (a_1, b) \leq_1 (a_2, b) \Rightarrow (a_1, b) \leq_2 (a_2, b) \wedge$$
$$(a, b_1) \leq_2 (a, b_2) \Rightarrow (a, b_2) \leq_1 (a, b_1)$$

$$(11)\ (a_1, b) \leq_1 (a_2, b) \Rightarrow (a_2, b) \leq_2 (a_1, b) \wedge$$
$$(a, b_1) \leq_2 (a, b_2) \Rightarrow (a, b_1) \leq_1 (a, b_2)$$

These conditions are symmetric if we swap the roles of players.

For one player a change of an action that (weakly) increases her payoff also (weakly) increases the payoff of the other player. However, for the other player a change of an action that (weakly) increases the payoff of the acting player (weakly) decreases first one's payoff.

One player is clearly a favourite of this game.

Human readable proof - example

Proposition 2

In games that satisfy (10) or (11) one player (the favourite) always has a strategy such that the best response of the other player is to "give" the favourite, her maximal payoff.

Intuition behind the proof:

1. Let's take an (a, b) with the maximal payoff for the favourite .
2. Either $\{b\} = B^{\star}(a)$ ✔ or there is some $b' \in B^{\star}(a)$.
3. By condition a change by the other player may not lead to a decrease in payoff for the favourite. Thus $(a, b') \simeq_2 (a, b)$ ✔.

Authors call this class of games "I-compete-you-cooperate".

Result

Similar results can be derived for other conditions. They make games either unfair in the sense mentioned above or not competitive (players has a profile that yield maximal payoff for both of them).

Conclusion

The class of *weakly unilaterally competitive games* is the most general class of “competitive” and “fair” games that have unique PNE payoffs.

Strict games

In the remaining part of the paper the same method is used to analyse strict games. Recall that a game is **strict** if for both players, different profiles have different payoffs. Moreover for strict games:

unique PNE payoff $\Rightarrow$ unique PNE (at most one PNE)

For strict games the only interesting uniqueness (sufficient) conditions that do not have dominant strategies are **weakly unilaterally competitive conditions** for individual player:

$$(a_1, b) \leq_1 (a_2, b) \Rightarrow (a_2, b) \leq_2 (a_1, b)$$
$$(a, b_1) \leq_2 (a, b_2) \Rightarrow (a, b_2) \leq_1 (a, b_1)$$

Strict games

Is there a sufficient and necessary condition for uniqueness of PNE?

Two games G_1 and G_2 are best-response equivalent iff for all $a \in A$: $B_1^\star(a) = B_2^\star(a)$ and for all $b \in B$: $A_1^\star(b) = A_2^\star(b)$

Theorem 3

A strict game has at most one PNE iff it is best-response equivalent to a strictly competitive game.

Intuition behind the proof: best response-equivalence preserves PNEs.

R.W. Rosenthal, Correlated equilibria in some classes of two-person games, International Journal of Game Theory (3) (1974) 119–128.

Conclusions

1. The automatic proof discovery may be used to find the weakest conditions – usually it is hard to prove that by hand.
2. It is hard to find unexpected or huge results in (part of) an intensively studied field.
3. Mixed strategies cannot be analysed in this framework.
4. Authors begin a new type of research in the field of game theory and indicate many open questions however there is no big follow-up. (Authors changed their interests, paper is cited mostly in overview papers and in Comsoc.)