Q-learning in Games: Collusion in Prisoner's Dilemma

Mentored By : Prof. Srinivas Arigapudi, Department of Economic Sciences, IIT Kanpur

Aayush Singh (220024)

Objective : When AI agents employ reinforcement learning algorithms to play classical games, do they converge to a Nash equilibrium? In this project, we attempt to answer the above question.

Approach:

- We study the pricing models as 2-player classical games where each of the players choose their strategy according to Q-learning (a reinforcement learning algorithm). We take the prisoner's dilemma, Hawk-Dove game, Coordination game and Bilingual Coordination game.
- Defined key parameters G and L and used an epsilon-greedy approach for strategy exploration.
- Conducted simulations for different values of the parameters to understand strategic dynamics.

Results:

1. Prisoner's Dilemma: If we take the values of discount factor to be zero they both are ending up in the D state while if we take into consideration the future payoffs ,they are going towards collusion
2. Hawk-Dove Game: for the smaller values of g they both are playing safe and are playing safe and choosing to go towards dove strategy ,for larger values of g(g>=0.3) one of them is ending up into hawk stage and other into dove stage respectively(for g belonging to 0 to 1)
3. Coordination game when the values of g is less than 1,both the players are playing safe and going towards state d and for the values of g greater than 1,depending on the initial states, they in some states are showing collusive behavior and going towards state c.
4. .Bilingual Coordination: As g increases, the payoff for mutual cooperation increases (i.e., both players choosing 'A').This often results in both players showing a preference for action 'A' when the cost c is relatively low. As c increases, the cost of miscommunication or cooperation decreases the attractiveness of these actions. Higher c values often result in players avoiding actions that involve cooperation with a cost, leading to a higher preference for the 'AB' action.

G,L>0

	Cooperate	Defect
Cooperate	1,1	-L,1+G
Defect	1+G,-L	0,0

G>0

	Hawk	Dove
Hawk	0, 0	1+G,1-G
Dove	1-G,1+G	1,1

0<G<1

	Option A	Option B
Option A	1+G,1+G	0, 0
Option B	0, 0	1,1

*Here G>0 assures (A,A) is payoff dominant Nash equilibrium, G<1 assures (B,B) is risk dominant Nash equilibrium

0<G<1,C>0

	A	B	AB
A	1+G,1+G	0,1	1+G,1+G-C
B	1,0	1,1	1,1-C
AB	1+G-C,1+G	1-C,1	1+G-C,1+G-C

C>0 implies that (AB,AB) is not a nash equilibrium)

Conclusion: Depending on the payoffs, and initial states, Q-learning algorithms may tend to converge towards collusive practices in different game scenarios, reflecting the potential collusion threats when implemented for AI-based pricing models.