Notes

Andrea Baisero

July 18, 2013

1 Rules

1.1 Food and Winning

Each player begins the game with $300(P-1)$ units of food, where P is the number of players.

If after any round you have zero food, you will die and no longer be allowed to compete. All players who survive until the end of the game will receive the survivor's prize.

The game can end in two ways. After a large number of rounds, there will be a small chance each additional round that the game ends. Alternatively, if there is only one person left with food then the game ends. In each case, the winner is the person who has the most food when the game ends.

The game is designed to encourage both cooperation and competition between players at different points in the game. Determining the right balance leads to a winning strategy.

1.2 Hunts

Each round is divided into hunts. A hunt is a game played between you and one other player. Each round you will have the opportunity to hunt with every other remaining player, so you will have $P-1$ hunts per round, where P is the number of remaining players.

In a hunt you can make the decision to hunt (become a hunter) or feign sickness and not hunt (become a slacker). Hunting is an active activity, so you will eat 6 units of food if you decide to hunt, and 2 units of food if you decide not to hunt. If both you and your partner do not hunt, then you get 0 food from the hunt. If only one of you decides to hunt then the hunt returns 6 units of food. If both of you decided to hunt then the hunt returns 12 units of food. The total food return of the hunt is split evenly between you and your partner. So, for example, if both of you decide to hunt, your individual net food gain/loss from the hunt would be $12/2-6=0$ units, whereas if both of you decided not to hunt your individual net food gain/loss is $0/2-2=-2$ food.

1.3 Reputation

Each player has a reputation R, which is defined as $R = \frac{H}{H+S}$, where H is the number of times you chose to hunt and S is the number of times you chose to slack (not hunt), both counted from the beginning of the game. Each player begins with a reputation of 0.

Each round, your tribe can save some time and hunt extra food if enough people opt to hunt. Each round, a random integer m, with $0 < m < P(P-1)$ will be chosen. If the total number of hunters is greater than or equal to m for the upcoming round, then everyone in the tribe gets $2(P-1)$ extra units of food after the round, which means that on average everyone gets 2 units of food per hunt. Before each round, you will find out the value of m.

Before each round you will be given: which number round you are in (from 1 to ∞), your food, your reputation, m, and an array with the reputation of each remaining member of the tribe when the round started. The location of each player in the array will be randomized after each round. You should decide your moves for that round for each player you will play against, and return an array of your decisions. After each hunt you will be given the food earned from each hunt you played.

1.4 Submission guidelines

The logic of your code and what game theory analyses you apply for the various steps must be either laid out as comments in the code or in a separate document. The algorithm (and separate document, if any) should be submitted by August 8, 2013 through your Brilliant account. Algorithms will be run tournament-style to determine the winner. The finalist teams will then have a short video interview with Brilliant staff, to verify that their work is their own.

Prizes for algorithms eligible for the final game will be awarded at two levels:

Grand Prize Winners – 5 teams will receive the Grand Prize of $1000. The winning algorithm is guaranteed to receive the Grand Prize. 4 other winners will be selected based on the performance of their algorithm, logic, game theory analysis, and presentation. Finalists – Members of teams whose algorithm survives the game will receive an "I survived" t-shirt.

2 Recap of Rules

Annotation

3 Strategy

3.1 Rewards

Reward Table

The reward table describes the reward functions W.

Symbol	Meaning
T	Current number of turn. Starts with 1.
P_t	Number of players at turn t.
R^i	Agent i's reputation. In the case of $i = 0$, then one can also index using turn number R^i_t.
h^i_t	Number of hunts the agent i performs at turn t. $i = 0$ denotes our agent.
H^i_t	Number of hunts from the beginning until turn t. $H^i_t = \sum_{\tau=1}^{T} h^i_\tau$. Also calculated as $H^i = \frac{R^i}{T}$.
S^i_t	Number of times agent i has slacked from the first turn. Calculated as?.
A^i	Agent i's action. $i = 0$ denotes my agent.
H_i	Number of times agent i has hunted. Calculated as $H_i = \frac{R_i}{N}$.
$W_0(A_0, A_i)$	Reward function for my agent. First action always indicates my action.
$W_1(A_0, A_i)$	Reward function for other agent. First action always indicates my action.

$W =$

$A_0 \setminus A_i$	H	S
H	$0 \setminus 0$	$-3 \setminus +1$
S	$+1 \setminus -3$	$-2 \setminus -2$

Expected Rewards

Let's try to analyse the expected rewards conditioned on the opponent's reputation.

$$\begin{aligned} E[W_0(A_0, A_i)/A_0 = H] &= E[W_0(H, A_i) \\ &= W_0(H, H)R_i + W_0(H, S)(1 - R_i) \\ &= 0R_i + -3(1 - R_i) \\ &= -3 + 3R_i \end{aligned}$$

$$\begin{aligned} E[W_0(A_0, A_i)/A_0 = S] &= E[W_0(S, A_i) \\ &= W_0(S, H)R_i + W_0(S, S)(1 - R_i) \\ &= 1R_i + -2(1 - R_i) \\ &= -2 + 3R_i \end{aligned}$$

Notice how the second is always higher than the first. This simply confirms that locally, the best action is to always Slack.

What if we also model our own action stochastically using our reputation?

Come to think of it, the following is not very useful.. We don't have uncertainty over our own action. I guess I was trying to model the opponent's action depending on our own reputation, but it is badly done in this way. I'll keep it just to keep track of all I've thought about, but ignore.

$$
\begin{aligned}
E[W_0(A_0, A_i)] &= E[W_0(A_0, A_i)/A_0 = H]R_0 + E[W_0(A_0, A_i)/A_0 = S](1 - R_0) \\
&= (-3 + 3R_i)R_0 + (-2 + 3R_i)(1 - R_0) \\
&= -2 - R_0 + 3R_i
\end{aligned}
$$

This result confirms the following: *a*) Whenever we chose to slack rather than hunt, we gain 1 reward point, *b*) Whenever the opponent chooses to hunt rather than to clask, we gain 3 reward points.

Conclusion

Choosing a policy which maximises the previous expectation leads to a static policy of always slacking, which overall would lead to a consistent loss of 2 items of food for every turn.

However, we will see that maybe we are not really using all the information we have. Up to now, we have assumed that the opponent's action is only dependent on his responsibility, but we have left out that his actions in front of *us* are very likely to depend on our own responsibility.

The next section will try to analyse why and how our reputation changes our opponents' actions, and how to make use of this to increase our expected food value.

3.2 Reputation Balancing

Why?

We have already seen that slacking is the Nash Equilibrium solution, i.e. the local optimum when we try to maximize the expected rewards. Unfortunately, this is far from a global optimum: Being this a prisoner dilemma problem, we have that both players gain by cooperating, but only if they both cooperate. This is the problem we are trying to address.

It is reasonable to assume that *some* players will try to cooperate so that they all gain by hunting together. If such a group forms, then I would rather be in it than out of it. The intended goal is to avoid being classified as a "slacker": If you are a slacker, then everyone will slack against you.

The problem is that there is no means of communication with the community so as to organize the up-coming of such a group: The only information we can exchange is our reputation. The goal of reputation balancing should be to send a message to your opponent: "look: you can trust me".

On the other hand the creation of such a group opens ground for all sorts of treason and backstabbing. Confirming yourself as a strong hunter is also not a good strategy: If you are a compulsive hunter then other people will assume they can slack their way into profit at your expenses. All in all, we just want to keep our reputation just high enough for that to the message to pass. We gain the maximum when we cooperate just enough to let the message pass. The question now is: *what is the threshold?*

One other interesting thing to consider is that the message we pass is only effective if there is someone listening on the other end. If we are surrounded by compulsive slackers, then there is little we can do but slack ourselves[1].

Idea: Would it be a sensible strategy to make our own reputation approach the global reputation? Possibly, but not always (e.g. if we are surrounded by strong hunters, or if we are head to head against just one last player and we have more food than him).

What value to aim for?

This is a very tough question.

Asssuming we would simply want to avoid the opponent to predict our own decision, then we would want to maximize the future entropy

$$H_f(R_{f,0}) = -R_{f,0}\log(R_{f,0}) - (1 - R_{f,0})\log(1 - R_{f,0})$$

To maximize, we simply aim at having $R_{f,0} = 0.5$, which is quite intuitive considering our goal of avoiding possible predictions by our opponent.

Given the current reputation R_0 and the number of total decisions we have to make P_t, then we can determine $R_{f,0}$ as a function of the times we are going to hunt in the current turn H_t.

$$R_{f,0} = \frac{H + H_t}{H + S + P_t}$$

How?

3.3 Rewards – reprise

In this section we will try to find maximise our expected gain given not only the opponent's reputation, but ours too, on the ground that both of them affect our certainty of what the opponent's action might be.

$$E[W_0(A_0, A_i)/R_0, R_i] =?$$

3.4 Public Goods

> "In many realistic situations there are benefits to cooperation that accrue to all players, but only if a certain number of players cooperate. These benefits are called public goods. A typical public good is a community playground built by donations. The playground can be built only if enough people help, but the playground benefits everyone equally. Public goods tend to increase the chance that people will cooperate."

[1] This is the reason why might want to keep a global reputation measure at hand, but more on that later.

from *Training Exercise*

How should the public goods in this game influence our actions? I can't find any way. By instinct, I would just ignore it.. But let's not be rash and think about it a bit more ☺.

3.5 Mixes Strategies

Another very interesting topic is that of mixed strategies, i.e. stochastic strategies in which instead of choosing an action, you chose what probability you assign to each action, and then execute the action stochastically.

3.6 Final Strategy

So how should one frame the decision-making?

- As a regression problem?
- As an optimization problem?