Intelligent Multi Agent Systems

TECHNISCHE UNIVERSITÄT DARMSTADT

Summer Semester 2015, Homework 1 (78 points + 11 bonus)
G. Neumann, G. Gebhardt, R. Lioutikov

Due date: Tuesday 26$^{\text{th}}$ May, 2015, 23:59 CEST (that's 10:59 AM on the following day in Pago Pago!)

How to hand in: Besides the Matlab code, the solutions have to be handed in on paper (either in <u>human readable</u> caligraphy or LaTeX) by putting them into the mailbox in front of S2|02 E315. The <u>well commented</u> Matlab code has to be handed in via moodle. Working in groups of up to <u>three people</u> are allowed, however every student needs to submit on its own (both, paper work and Matlab code) and name his collaborators.

Problem 1.1 Theoretical Questions

Keep your answers short! As a rule of thumb: not more than two sentences per point!

a) **Types of Interaction** **[3 Points]**
Name and explain the three types of interaction and give a real world example for each.

Cooperatively: Solving a task together, e.g. Homework in a group.
Neutrally: Solving non-competitive task independently, e.g. Driving a car inner city.
Competitively: Solving opposing task, e.g. Football (soccer).

b) **Rational, Autonomous Agent** **[2 Points]**
Give a short description of a rational, autonomous agent.

A **rational** agents optimizes it's performance measure. A **autonomously** agent uses it's own perception to regulate itself.

c) **Interaction** **[3 Points]**
Name the important quantities for the interaction of an agent with its environment and give a short description for each.

The agent acts upon the world state with a given sets A_i of **actions**. Each action is assigned with an **utility** u_i. Interaction means *communication*. An agents communication can be quantified via the spoken **language** and **protocols**.

d) **Reflex Policy** **[2 Points]**
What is a reflex policy and how does it differ from a normal policy?

A normal policy includes the observation *history* $\pi(o_{1:t}) = a_t$. A **reflex policy** takes only the current observation into account $\pi(o_t) = a_t$.

e) **Markov Property** [2 Points]

How does the Markov property influence the policy?

The **markov property** uses the bayesian property to explain, that the current state of an environment and its participants, also describes it's action and state *history* $p\left(s_{t+1}|s_{a:t}, a_{1:t}\right) = p\left(s_{t+1}|s_t, a_t\right)$.

f) **Observability** [3 Points]

Give a definition for an *observable* and a *partially observable* world. Assuming Markovian dynamics, how does this influence the policy?

An **observable world** provides the agent with *complete information* about the world $o_t = s_t$. A **partially observable** world provides the agent with information drawn from a distribution $o_t \sim p\left(o_t|s_t\right)$.

g) **Utility Function** [1 Points]

What is a utility function?

The utility function is "a mapping from states of the world to real numbers, indicating the agent's level of happiness with that state of the world ". *From: 01-Introduction slide 24*

h) **Best Response Operator** [2 Points]

Give a written definition of the best response operator (no mathematical definition).

The **best response operator** B_i for player i chooses the action a_i, that maximizes the players i utility u_i, given the knowledge of how every other players will act a_{-i}.

i) **Nash Equilibria** [1 Points]

Are there always pure strategy/mixed strategy Nash equilibria?

...or correlated *Nash equilibria*, Yes. (At least in the definition of game theory, we know so far.)

j) **Zero-Sum Games** [2 Points]

What are the characteristics of a zero-sum game? Describe them briefly.

[Zero-sum.] The sum of the players utilities for any single time step t is zero: $\sum_i \left(u\right) = 0$.
[Competition.] Since the utilities sum of the agents actions is zero, the agents are in competition to each other.
[Nash equilibria.] Since there has to be a *looser* for any time step t, the Nash equilibrium has to be a *mixed Nash equilibrium*.

Name, Vorname: _____   Matrikelnummer: ⌴⌴⌴⌴⌴⌴⌴⌴

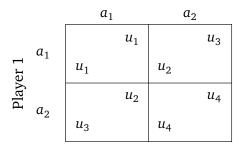---

Problem 1.2 Pen & Paper Exercises

a) **Games**   **[15 Points]**
Consider Pay-Off Table 1:

Pay-Off Table 1

Player 2

	a_1	a_2
a_1	u_1 / u_1	u_3 / u_2
a_2	u_2 / u_3	u_4 / u_4

(Player 1 on the left)

1. Which game does Pay-Off Table 1 describe for the order

$$u_3 > u_1 > u_4 > u_2?$$

   This game describes the Prisoners Dilemma

2. Give the order of the utilities for the stag hunt game.
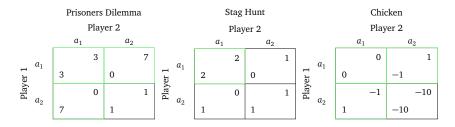
$$u_1 > u_3 = u_4 > u_2?$$

3. Give the order of the utilities for the hawk-dove (chicken) game.

$$u_3 > u_1 > u_2 > u_4?$$

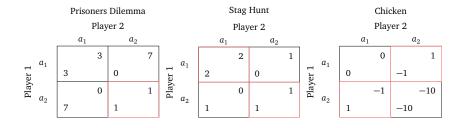4. Which of these games are coordination games?

   The Stag-Hunt game is a coordination game, seeing as both players can maximize their reward if they work together.

5. Give the Pareto optimal solutions for all three games in separate tables (in cornflower blue).

Prisoners Dilemma
Player 2

	a_1	a_2
a_1	3 / 3	7 / 0
a_2	0 / 7	1 / 1

Stag Hunt
Player 2

	a_1	a_2
a_1	2 / 2	1 / 0
a_2	0 / 1	1 / 1

Chicken
Player 2

	a_1	a_2
a_1	0 / 0	1 / −1
a_2	−1 / 1	−10 / −10

6. Give the pure strategy Nash equilibria for all three games (in brick red).

Prisoners Dilemma

Player 2

	a_1	a_2
a_1	3 / 3	7 / 0
a_2	0 / 7	1 / 1

Stag Hunt

Player 2

	a_1	a_2
a_1	2 / 2	1 / 0
a_2	0 / 1	1 / 1

Chicken

Player 2

	a_1	a_2
a_1	0 / 0	1 / −1
a_2	−1 / 1	−10 / −10

7. Fill Pay-Off Table 2 such that action $a_{1,1}$ stricktly dominates action $a_{1,2}$ and such that action $a_{2,1}$ weakly dominates action $a_{2,2}$.

Pay-Off Table 2

Player 2

	a_1	a_2
a_1	5 / 5	4 / 4
a_2	3 / 3	2 / 3

b) **Battle of Food**   **[12 Points]**

It's caketime, the stomach growls, Chris and Herke are hungry. Chris prefers Mettbroetchen from the butcher, Herke prefers cake from the bakery. When Chris eats a Mettbroetchen his dopamine level rises by 3 units, if he eats cake it rises by only 2 units. A piece of cake rises Herke's dopamine level by 3 units and a Mettbroetchen only by 2. Additionally, going for a snack together doubles the release of dopamine. However, if Chris gets himself a cake and knows that Herke has the joy of a Mettbroetchen in the meantime, he gets so upset that he throws away the cake.

1. Draw the pay-off table.

Pay-Off Table

Herke

	Mb	Ck
Mb	6 / 4	2 / 3
Ck	0 / 2	4 / 6

2. Mark (in magenta) and explain the Pareto optimal outcomes.

Name, Vorname: _____   Matrikelnummer: ⌐⌐⌐⌐⌐⌐⌐⌐

Pay-Off Table

Herke

		Mb	Ck
		4	3
Mb	6		2	
		2	6
Ck	0		4	

(Chris, rows)

In this game the pareto optimality lies where both people select the same food. By definition pareto optimality means that we cannot improve on one person without harming the other, and in this case we can see that either of the outcomes where they pick the same food the "social welfare" is ideal. to start with the different choice scenario has lower values for either of them, so there is no point in switching to those lesser values from either person point of view. The other option is to switch from the lesser value(4) to the other maximum (6), but that would require lowering the other persons value, thus breaking the description.

3. Mark (in azur) and explain the pure Nash equilibria.

Pay-Off Table

Herke

		Mb	Ck
		4	3
Mb	6		2	
		2	6
Ck	0		4	

(Chris, rows)

Pure nash Equilibria is actually the same as the pareto optimality, mostly because if either person already knows what the other will choose then it makes sense to try and achieve the highest posible value for them (4 in the worst case, 6 in the other) by choosing to eat together. The reward for eating their favorite food alone is not big enough to outvalue the sharing scenario. This can be seen in every day of our lives, when we go out to eat with friends there might be people who prefer eating chinese or mexican food, but eating together is way better than eating alone thus we compromise.

4. Compute the mixed strategy Nash equilibria (show the steps).

Chris pure strategy Mettbrotchen value : $6q + 2(1-q) = 4q + 2$
Chris pure strategy Cake value : $4(1-q) = -4q + 4$
Herke pure strategy Mettbrotchen value : $4p + 2(1-p) = 2p + 2$
Herke pure strategy Cake value : $3p + 1(1-p) = -3p + 6$

Chris:
$4q + 2 = 4q + 4$ $\qquad$ $8q = 2$ $\qquad$ $q = 0.25$
$4q + 2 < 4q + 4$ when $q < 0.25$
$4q + 2 > 4q + 4$ when $q > 0.25$
$4q + 2 = 4q + 4$ when $q = 0.25$

Herke:
$2p + 2 = -3p + 6$ $\qquad$ $5p = 4$ $\qquad$ $p = 0.80$
$2p + 2 < -3p + 6$ when $p < 0.80$
$2p + 2 > -3p + 6$ when $p > 0.80$
$2p + 2 = -3p + 6$ when $p = 0.80$

The system is stable under the best response q=0.25 and p= 0.8 with utility (3,3.6).

c) **The Dancing Queen** $\hfill$ **[10 Points]**
   ♪Alice is the girl that Eve's been dreaming of ever since she was a little girl. ♫

   Eve has a crush on Alice, but Alice is, unfortunately, together with Bob. When the three of them go clubbing together, Eve lives in a constant up and down of feelings. On the one hand, Bob is a terrible dancer, so Eve has to suffer the pain of watching Bob stepping on Alice's flawless feet. On the other hand, Bob is a terrible dancer, so Alice dances every second song with Eve, which gives Eve a reward of 100 golden coins. But Eve also knows that Bob is not the most loyal guy, so there is a probability of 10% that he will cheat on Alice and the two will break up. This would be worth an additional 100 golden coins for Eve. As Eve is a state-approved dance instructor, she could also teach Bob how to dance with Alice. Though, since Eve doesn't like Bob, that would cost her 20 golden coins. However, as a perfect dancer, Bob would attract much more women and, thus, the probability of Bob breaking up with Alice would rise to 60%. At the same time, Alice would enjoy the dances with Bob, the perfect dancer, more and hence would dance only every fifth song with Eve.

   1. Draw the utility table.

		Don't Teach Bob Dance (0.5)				Teach Bob Dance (0.2)	
		Y	N			Y	N
Breakup((0.1))	Y	200	100	Breakup((0.6))	Y	180	80
	N	100	0		N	80	−20

2. Compute the expected utility of Eve for teaching Bob how to dance (show the steps).

   $u(teach) = 180pq + 80(1-p)q + 80p(1-q) + -20(1-p)(1-q)$
   $u(teach) = 180(0.2)(0.6) + 80(0.8)(0.6) + 80(0.2)(0.4) + -20(0.8)(0.4)$
   $u(teach) = 21.6 + 38.4 + 6.4 - 6.4$
   $u(teach) = 60$

3. Should Eve teach Alice's boyfriend how to dance with her?

   It actually does not matter in the least, because the result of the utility function for the case where she does not teach bob how to dance gives the same utility of 60, so the results are the same in either case, and should spend her time with something more useful. Or just find another girl without boyfriend.

4. Sketch your favorite dance moves **[0 Points, but a smiley]**

Problem 1.3 Matlab Exercises

a) **Rock, Paper, Scissors, etc.**   **[20 Points]**
   Download the Matlab templates from moodle. Implement the maxmin and minmax algorithms as well as their variations with slack variables in the respective templates. Do not modify the signature of the functions. Answer the following questions on paper.

   - Which strategy do you obtain with the minmax routine, which one with the maxmin routine?

   - Rock, Paper, Scissors[1]

     1. Write down the payoff table for the game *rock, paper, scissors*.

     2. Which kind of game is rock, paper, scissors?

     3. What strategy do you obtain, when feeding the payoff table of player 1 into your implemented routines?

     4. What are the slack variables you obtain and what does this mean for the available actions?

     5. Assume that player 2 only has the actions rock and paper, what are the optimal strategies and values now for both players? What are the slack variables you get? Is this still the same kind of game as the normal rock, paper, scissors game?

   - Rock, Paper, Scissors, Well[2]

---

[1] http://en.wikipedia.org/w/index.php?title=Rock-paper-scissors
[2] http://en.wikipedia.org/w/index.php?title=Rock-paper-scissors#Additional_weapons

1. Write down the payoff table for the game *rock, paper, scissors, well*.

2. Is this still the same kind of game as rock, paper, scissors?

3. What strategy do you obtain, when feeding the payoff table of player 1 into your implemented routines?

4. What are the slack variables you obtain and what does this mean for the available actions?

- Rock, Paper, Scissors, Lizard, Spock[3]

1. Write down the payoff table for the game *rock, paper, scissors, lizard, spock*.

2. Is this still the same kind of game as rock, paper, scissors?

3. What strategy do you obtain, when feeding the payoff table of player 1 into your implemented routines?

4. What are the slack variables you obtain and what does this mean for the available actions?

b) **LCP Baby!**   **[11 Bonus Points]**

Implement the function `lcp_nash`. The function takes a tensor with the payoff tables for both players and the initial values for the variables as arguments and returns the utilities, the strategies and the slack variables for both players.

Test your implementation with the *battle of sexes* game from the lecture. Try to find all the Nash equilibria by trying different initializations of the variables. Which Nash equilibria do you get for which initializations?

---

[3] http://en.wikipedia.org/w/index.php?title=Rock-paper-scissors#Additional_weapons