CSCE 475H Multiagent Systems

Examination - Solution

March 31, 2016

Name:

NUID:

All students must answer the REQUIRED questions (#1 and #5).

Choose ***three*** (3) out of the remaining four questions to respond to. If you attempt more than three, your ***must*** indicate which three you want to be graded.

You have 75 minutes to complete the examination.

1. (REQUIRED) (20 total points) Agency.

(a) (4 points) What is an agent?

Solution: An agent is an entity that senses its environment, makes autonomous decisions based on the sensory input, and actuates the decisions that in turn change the environment.

(b) (3 points) What is an intelligent agent?

Solution: An intelligent agent is an agent capable of flexible behavior: reactive (responsive in a timely manner), proactive (goal-directed) and social. Another definition calls for an agent to be capable of learning in order to be intelligent.

(c) (3 points) What is a multiagent system?

Solution: A multiagent system is an environment where there are multiple agents interacting directly or indirectly such that the system designer obtains an overall system result, usually one that cannot be achieved with each individual agent alone.

(d) (10 points) Under which ***five*** environment characteristics is it more appropriate to use an **agent-based** solution? Identify a problem and describe its environment with respect to these five characteristics.

Solution: Dynamic (environment changes while decisions are being made), Partially observable (or incomplete information), Uncertain (or stochastic such that the same action does not lead to the same outcome), Non-episodic (that state values depend on past history), and Continuous (where the number of states is infinite). When the environment is dynamic, or partially observable, or uncertain, or non-episodic, or continuous, it is more appropriate to use an agent-based solution so that the autonomy and decentralized reasoning can address these complex characteristics better. A smart grid simulation problem: dynamic, partially observable, uncertain, probably episodic, and the states can be quite discrete (e.g., breaking a day's electricity usage to different phases: after getting up, away for work, back from work, sleep).

2. **(20 points) Learning. Definition 7.4.1 (Q-learning)** *Q-learning is the following procedure:*

Initialize the Q-function and V values (arbitrarily, for example)
repeat *until convergence*
Observe the current state s_t.
Select action a_t and take it.
Observe the reward $r(s_t, a_t)$ and next state s_{t+1}
Perform the following updates (and do not update any other Q-values):

$$Q_{t+1}(s_t, a_t) \leftarrow (1 - \alpha_t)Q_t(s_t, a_t) + \alpha_t(r(s_t, a_t) + \beta V_t(s_{t+1}))$$

$$V_{t+1}(s) \leftarrow \max_a Q_t(s, a)$$

(a) (6 points) Explain the purposes of the factors α_t and β.

Solution: α_t is the learning rate. It is used to control the amount of "trust" in the knowledge that an agent has learned so far. If the value is low, then it means the agent will tend to adhere to what it has learned before. If the value is high, then the agent will be more willing to weigh current rewards more. β is the discount factor on the future "expected" value. A higher value of β means the agent is more willing to lookahead into the future, and vice versa. In other words, an agent with a higher learning rate will tend to explore more (more adaptive to the environment), and an agent while an agent with a lower discount factor will tend to be more myopic.

(b) (8 points) Explain the intuition behind the update for the Q value, especially the roles of the three terms, $(1 - \alpha_t)Q_t(s_t, a_t)$, $\alpha_t(r(s_t, a_t))$, and $\alpha_t \beta V_t(s_{t+1})$, in the equation.

Solution: The intuition behind the update for the Q value is to combine what an agent has learned before, what it has just received as reward to its action, and then what it projects that it would receive in the future from being in the resulting state. The value of each state is also updated as the maximum of the Q value that the state is able to garner over all possible actions.

The first term is about the existing knowledge, allowing an agent to "reuse" what it has learned. The second term is the immediate reward to an action that an agent has done while in a state, allowing an agent to focus and exploit the actual reward received. The third term is the value of the resulting state, allowing an agent to lookahead in the future when considering the Q-value of a state-action pair.

(c) (6 points) Situation #1: If the MAS environment is very uncertain, what should β be? Justify your answer. Situation #2: If agents in the environment will only encounter a few learning episodes, what should α_t be? Justify your answer.

Solution: Situation #1. β should be low since the future values cannot be trusted due to high uncertainty. Situation #2. With only a few opportunities, an agent can try to be more risk-seeking with a high α_t to learn fast and adapt.

3. (20 total points) Voting. Given the following Definition (Groves mechanisms)
Groves mechanisms *are direct quasilinear mechanisms* $(\chi, \wp)$, *for which*

$$\chi(\hat{v}) = arg \max_x \sum_i \hat{v}_i(x),$$

$$\wp_i(\hat{v}) = h_i(\hat{v}_{-i}) - \sum_{j \neq i} \hat{v}_j(\chi(\hat{v})).$$

(a) (6 points) What does $\chi(\hat{v})$ do? What does $\wp_i(\hat{v})$ do? What does each of the components in $\wp_i(\hat{v})$ do?

Solution: $\chi(\hat{v})$ is the social choice function; i.e., it finds the best choice that gives the maximum total agent valuations. $\wp_i(\hat{v})$ is the payment function, that an agent *i* has to pay the voting system. It says that an agent *i* is responsible to pay a "penalty" or "tax" based on a heuristic function that takes the valuation without *i*'s declaration. However, an agent can pay less as this penalty is subtracted by $\sum_{j \neq i} \hat{v}_j(\chi(\hat{v}))$, which is the total agent (without *i*) valuations of the social choice. That is, an agent gets to pay less if other agents value the social choice more.

(b) (6 points) What is the dominant strategy under any Groves mechanism? Justify your answer.

Solution: The dominant strategy under any Groves mechanism is truth-telling. This is because (1) if an agent under-values a candidate of its choice, then it is possible that the candidate might not be chosen, and thus the agent would not be able to gain the utility of having that candidate chosen, and (2) if an agent over-values, then if the candidate is selected, then other agents get to pay less, but itself will not pay less. Thus, an agent will have no motivation but to reveal its true valuation. This is because the reward to an agent relies not on how much it could gain from manipulating its own valuation, but on how much it could help the rest of the agents benefit from the social choice function outcome.

(c) (8 points) What is the Clarke Tax definition for $h_i(\hat{v}_{-i})$? What is the rationale? (***Hint***: Pivotal agents)

Solution: The Clarke's tax is $h_i(\hat{v}_{-i}) = \sum_{j \neq i} \hat{v}_j(\chi(\hat{v}_{-i}))$. Assume that all agents follow their dominant strategies and declare their valuations truthfully. The second sum in the VCG payment rule pays each agent i the sum of every other agent $j \neq i$'s utility for the mechanism's choice. The Clarke's tax, however, charges each agent i the sum of every other agent's utility for the choice that *would have been made* had i not participated in the mechanism. Thus, each agent is made to pay its *social cost*—the aggregate impact that its participation has on other agents' utilities.

If some agent i does not change the mechanism's choice by its participation—that is, if $\chi(v) = \chi(v_{-i})$—then the two sums in the VCG payment function will cancel out. The social cost of i's participation is zero, and so it has to pay nothing.

In order for an agent i to be made to pay a nonzero amount, it must be *pivotal* in the sense that the mechanism's choice $\chi(v)$ is different from its choice without i, $\chi(v_{-i})$. This is why VCG is sometimes called the ***pivot*** mechanism—***only pivotal agents are made to pay***.

4. **(20 points) Auction**. Consider a *cartel* or a *bidding ring*, with the Vickrey auction protocol.
(a) (10 points). Describe the protocol.

Solution: The protocol as follows.

1. Each agent in the cartel submits a bid to the ring center.
2. The ring center
   a. identifies the maximum bid that it received, $\hat{v}_1^r$;
   b. submits this bid in the main auction; and
   c. drops the other bids.

   (Denote the highest dropped bid as $\hat{v}_2^r$.)
3. If the ring center's bid wins in the main auction (at the second-highest price in that auction, $\hat{v}_2$), the ring center awards the good to the bidder who placed the maximum bid in the cartel and requires that bidder to pay $max(\hat{v}_2, \hat{v}_2^r)$.
4. The ring center gives every agent who participated in the bidding ring a payment of k, regardless of the amount of that agent's bid and regardless of whether or not the cartel's bid won the good in the main auction.

(b) (10 points). What are the situations where the ring center (leader) would lose money for running the ring? What are the situations where the ring center (leader) would gain?

.

Solutions: If $\hat{v}_2^r > \hat{v}_2$, the ring leader will pay $\hat{v}_2$ for the good in the main auction, but it will be paid $\hat{v}_2^r$ for it by the winning bidder. Let $c > 0$ denote the ring center's expected profit. If there are n_r agents in the ring, then the ring center could pay each agent up to $k = \frac{c}{n_r}$ and still budget balance on expectation! For values of k smaller than this amount but greater than zero, the ring center will profit on expectation while still giving agents a strict preference for participation in the bidding ring. This means that if the ring has high bidders than the main auction, then the ring leader stands a good chance to gain; if the main auction is likely to have the high bidders and thus also winning bidders, then the ring leader stands to lose as it now needs to still pay ring members k.

5. **(REQUIRED) (20 points) Problem Solving with MAS.**
(a) (10 points) *Local autonomy vs. globally directed solutions*
(i) Why is local autonomy desirable?

Solution: We want to use an agent-based solution because we want the agent autonomy—making local decisions on their own—so that the solution can be responsive, robust, adaptive, and so forth.

(ii) What is one key concern regarding using a system of fully autonomous agents?

Solution: However, with agents making their local decisions, that means there is not a centralized control. In general, without a centralized control, it is not guaranteed that the system will exhibit desired, coherent behaviors.

(iii) Why are globally directed solutions desirable?

Solution: System designers impose global or top-down directives or rules to guarantee solutions.

(iv) What is one key concern regarding using a system with globally directed solutions?

Solution: Using a system with globally directed solutions reduces agent autonomy, hereby defeating the purpose of wanting an agent-based solution in the first place, and losing the benefits of adaptiveness, responsiveness, scalability, robustness, and so forth.

(v) What is the elegance of emergent behavior in a MAS? Why is this idea central to a good MAS design?

Solution: The elegance of emergent behavior in a MAS is that the desirable overall behavior is not pre-programmed into the agents' reasoning, and thus restricting their decision making and imposing a certain set of solutions on them. This is central to a good MAS design as it allows each agent to make decision locally with as much autonomy as it needs, and together, the agents' joint actions still lead to desired, emergent coherent behavior overall.

(b) (10 points) Describe your final project and explain how your experiment/study aims to investigate the above in your project.

Solution: This of course has many different answers. In general, one should provide the agent local decisions, the desired emergent behaviors are, and the experiment setup and hypotheses.

6. (20 points) Hodgepodge.

(a) (6 points) What is a "spoiler" in voting? Provide also an example using the plurality voting scheme that shows a spoiler.

Solution: Consider the following preferences by 100 agents.

$$35\ agents: a \succ c \succ b$$
$$33\ agents: b \succ a \succ c$$
$$32\ agents: c \succ b \succ a$$

Plurality would pick candidate a as the winner. However, if the candidate c did not exist, then plurality would pick b. So, candidate c is a spoiler. A spoiler is a candidate that stands no chance of being selected but could change the selected outcome.

(b) (7 points) Setup the Stackelberg game. Further, the game has a Nash equilibrium. Which one? Explain. One of the players has a dominant strategy. Which one? Explain.

Solution: The Stackelberg game:

	Player 2 plays L	*Player 2 plays R*
Player 1 plays T	1,0	3,2
Player 1 plays B	2,1	4,0

The Nash equilibrium is when Player 1 plays B and when Player 2 plays L. When Player 2 players L, Player 1 can either play T or B. But T gives Player 1 only 1 utility point. So, Player 1 will choose to play B. When Player 1 plays B, Player 2 can either play L or R. But, R gives Player 2 only 0 utility point. So Player 2 will choose to play L. Thus, this is the equilibrium as neither player has the motivation to change how they play. Player 1 has a dominant strategy. This is because regardless of how Player 2 plays, playing B always gives Player 1 a better utility (2 vs. 1 when Player 2 plays L, 4 vs. 3 when Player 2 plays R).

(c) (7 points) Given the table of preferences from five different voters for their favorite states in the Midwest Region, using the Borda count scheme and pairwise elimination order, which state is the winner? (Please fill out the Table below.)

States	Voter 1	Voter 2	Voter 3	Voter 4	Voter 5	Total Borda Counts
Colorado	7	0	7	5	4	23
Illinois	6	1	0	6	5	18
Indiana	5	2	1	7	6	21
Iowa	4	3	2	0	7	16
Kansas	3	4	3	1	0	11
Missouri	2	5	4	2	1	14
Nebraska	1	6	5	3	2	17
Oklahoma	0	7	6	4	3	20

Round 1	Round 2	Round 3	Round 4
Colorado			
Illinois			
Indiana			
Iowa			
Kansas			
Missouri			
Nebraska			
Oklahoma			

Solution:

Round 1	Round 2	Round 3	Round 4

Colorado			
Illinois	Illinois (3 vs 2)		
Indiana			
Iowa	Iowa (3 vs 2)	Iowa (3 vs 2)	
Kansas			
Missouri	Missouri (4 vs 1)		
Nebraska			
Oklahoma	Oklahoma (4 vs 1)	Oklahoma (4 vs 1)	Oklahoma (3 vs 2)

The winner is Oklahoma.