COMP 307/420 — *Introduction to AI*

Assignment 4: Planning and Scheduling

25% of Final Mark — Due: 23:59 Tuesday 23 June 2020

1 Objectives

The goal of this assignment is to help you understand the basic concepts of planning and scheduling, and simple algorithms for them. In addition, your understanding on search techniques, machine learning basics, and some other hot AI topics will be tested. In particular, the following topics should be reviewed:

- Use Planning Domain Definition Language (PDDL) to represent a classical planning problem,
- Use state-space search algorithms to find a plan,
- Represent a schedule of job shop scheduling problem,
- Generate a schedule by a dispatching rule for a given job shop scheduling problem,
- Review concepts on machine learning basics (lectures 2–6), and
- Understand broad knowledge of some hot topics in AI (Lectures 20-21).

All the questions in this assignment can be answered without programming. You can simply write the answers to the questions in your report.

2 Question Description

Part 1: Classical Planning – Monkey-and-Bananas [35 marks]

In this part, you are required to represent the monkey-and-bananas problem using PDDL, and find a plan by state-space search.

Problem Description

In the monkey-and-bananas problem, a monkey, a box and some bananas are placed in a room. The bananas are hung from the ceiling and the monkey cannot reach them directly. The only way for the monkey to catch the bananas is to move the box under the bananas and climb up onto the box.

Assume that initially, the monkey is at location A, the box at location B and the bananas at location C. The monkey and the box have height Low, i.e. they are on the ground. The bananas have height $High$. If the monkey climbs onto (down) the box, its height will be $High$ (Low). The monkey cannot catch the bananas when its height is Low. The actions that the monkey can take include (1) *Go* from one place to another, (2) *Push* the box from one place to another (assume this action requires that the box and the monkey are at the same current location, and as a result the box and the monkey are at the same new location), (3) *ClimbUp* onto or *ClimbDown* from the box (the box and the monkey are at the same location before and after these actions), and (4) *Grasp* and *Ungrasp* the bananas. The result of a *Grasp* is that the monkey holds the bananas, if the monkey and the bananas are in the same place at the same height. The goal state is that the monkey holds a banana.

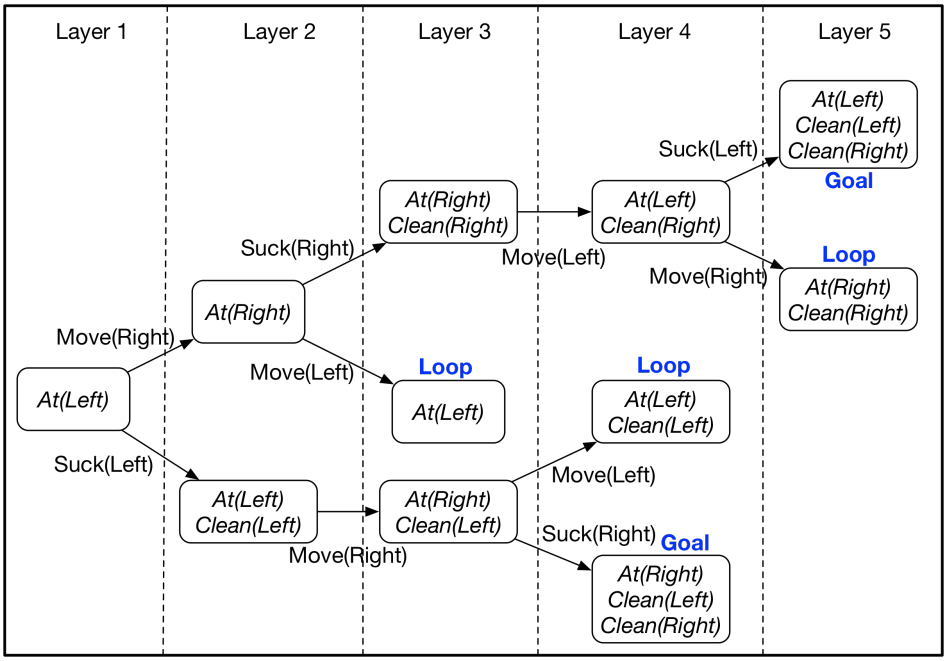
Questions

1. (5 marks) Write the description for the initial and goal states using PDDL.
2. (10 marks) Write all the action PDDL schemas (there are 6 actions in total). Each action should include a *name*, a set of *variables*, a *precondition* and an *effect*. The precondition of each action should be set in a way that self-connection is avoided, i.e. no action connects any state to itself.

3. (15 marks) A plan to achieve the goal state from the initial state can be found by using forward state-space search. Based on the PDDL in Questions 1 and 2, draw the **first three layers** of the corresponding state-space search graph to demonstrate the search process.

   - Each node (state) in the graph is represented by a conjunction of fluents.
   - Each edge is associated with an action.
   - Each leaf node is either a goal state or a loop.

   Example: The figure below is the 5-layer forward state-space search graph of the vacuum cleaner's world (the $\wedge$ between fluents are omitted).



4. (5 marks) Write a plan to achieve the goal state from the initial state. The plan needs to be formatted as follows:

   - Initial state:
   - Action 1:
   - State 1:
   - Action 2:
   - ...
   - State k (goal state):

5. **(for COMP420 ONLY, 10 marks)** Draw the **last three layers** of the backward (regression) state-space search.

Part 2: Job Shop Scheduling [40 marks]

In this part, you are required to find solutions (schedules) for a job shop scheduling problem.

Problem Description

The table below gives a job shop schedule problem with 3 jobs and 2 machines.

Job	ArrivalTime	Operation	Machine	ProcTime
J_1	0	O_{11}	M_1	50
		O_{12}	M_2	25
J_2	10	O_{21}	M_2	30
		O_{22}	M_1	35
J_3	20	O_{31}	M_1	40
		O_{32}	M_2	20

- (Number of operations) Each job J_j has two operations O_{j1} and O_{j2}.
- (Order constraint) The operations strictly follow the order constraint. That is, O_{j2} $(j = 1, 2, 3)$ cannot be processed until O_{j1} has been completely processed.
- (Arrival time) Each job has an arrival time (ArrivalTime). For each job J_j, the first operation O_{j1} cannot be processed earlier than its arrival time.
- (Resource constraint) Each operation can only be process by a particular machine. For example, operation O_{11} can only be processed by machine M_1. Each machine can process at most one operation at a time.

Solution/Schedule Representation

A solution/schedule for a job shop scheduling problem is a sequence of actions. Each action is composed of the processed operation, the machine to process the operation, and the starting time. The finishing time of an action is the starting time plus the processing time of the processed operation. The actions are sorted in the increasing order of starting time, i.e. the former action starts no later than the latter one. In this assignment, the following format is adopted to represent a schedule:

$$Process(O_{11}, M_1, 0) \rightarrow Process(O_{21}, M_2, 10) \rightarrow \ldots,$$

where $Process(o, m, t)$ stands for an action that processes the operation o with machine m and starts at time t.

Questions

1. (10 marks) Given a schedule whose action sequence is as follows: $Process(O_{11}, M_1, t_1) \rightarrow Process(O_{21}, M_2, t_2) \rightarrow Process(O_{31}, M_1, t_3) \rightarrow Process(O_{12}, M_2, t_4) \rightarrow Process(O_{22}, M_1, t_5) \rightarrow Process(O_{32}, M_2, t_6)$. Since the sequence is sorted in the non-decreasing order of starting time, we know that $t_1 \leq t_2 \leq t_3 \leq t_4 \leq t_5 \leq t_6$. Calculate the **earliest starting time** (t_1 to t_6) of each action. You can draw a gantt chart to help you think.

   Hint: the earliest starting time of an action is the later time between the earliest ready time of the operation and the earliest idle time of the machine.

2. (5 marks) For the solution given in Question 1, find the **completion time** of each job, which is the finishing time of its last operation. Then, calculate the **makespan** of the solution, which is defined as the maximum completion time of all the jobs.

3. (15 marks) Write the state from **step 1 to step 3**, and the **final solution** when applying the Shortest Processing time (SPT) dispatching rule to the problem. At each step, the representation of a state is composed of (1) a **partial solution**, (2) the **earliest idle time** of each machine and (3) the **earliest ready time** of each unprocessed operation. The initial state (step 0) is given below for your reference.

   Step 0:

   - Partial solution: (empty, no action is scheduled)
   - $earliestIdleTime(M_1) = 0$, $earliestIdleTime(M_2) = 0$
   - $earliestReadyTime(O_{11}) = 0$, $earliestReadyTime(O_{12}) = \infty$
   - $earliestReadyTime(O_{21}) = 10$, $earliestReadyTime(O_{22}) = \infty$
   - $earliestReadyTime(O_{31}) = 20$, $earliestReadyTime(O_{32}) = \infty$

4. (5 marks) For the solution obtained by the SPT rule, calculate the completion time of each job and the makespan. Compare the makespan between this solution with that obtained in Question 1 to find out which solution is better in terms of makespan.

   Note: the solution in Question 1 is obtained by the First-Come-First-Serve (FCFS) rule.

5. (5 marks) The two compared solutions are obtained by the SPT and FCFS rules, respectively. If one solution is better then the other, does it mean that the rule that generates the better solution is better than the other rule? Why or why not?

6. **(for COMP420 ONLY, 20 marks)** Often in practice, neither the SPT nor FCFS rules are good enough for solving the job shop scheduling problem. Suggest two methods to solve the job shop scheduling problem. One method should consider the problem to be static (all information is known in advance), and the other method should consider it to be dynamic (e.g. unpredicted job arrivals can happen in real time). **Clearly describe your methods (e.g. algorithm framework, solution representation/encoding, search operators).**

Part 3: Search Techniques and Machine Learning Basics: Questions [30 marks]

Question 1. (10 marks)

1. (3 marks) We have discussed a range of search methods in the lectures. For the following situations, state which search method is best to use, and briefly explain why.

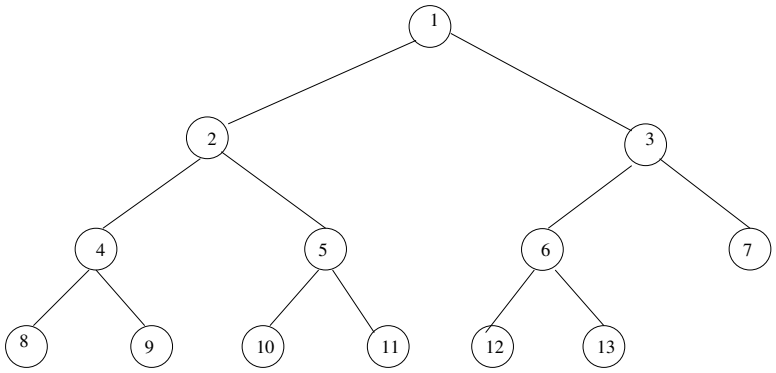
   (i) If only the cost/distance from the initial state to a particular current/intermediate state is known, and the cost/distance between a current/intermediate state and the goal state is not known.

   (ii) If only the cost/distance between a particular current/intermediate state and the goal state is known, and the cost/distance from the initial state to a current/intermediate state is not known.

   (iii) If both the cost/distance from the initial state to a particular current/intermediate state and the cost/distance between a current/intermediate state and the goal state are available.

2. (2 marks) Simon is trying to solve a scheduling problem for a manufacturing company. He tried both the hill climbing and genetic beam search methods, and found that genetic beam search obtained much better solutions than the hill climbing method. Briefly explain a possible reason.

3. (5 marks) Consider the tree below, state the search order/path using the numbers in the nodes for (i) *breadth-first search* and (ii) *iterative deepening search*.



Question 2. (10 marks)

Michael aims to develop a classification model based on the following 10 training instances from the Iris dataset.

Sepal Length (mm)	Sepal Width (cm)	Patel Length (cm)	Patel Width (cm)	Class
51	3.5	1.4	0.2	setosa
49	3.0	1.4	0.2	setosa
47	3.2	1.3	0.2	setosa
70	3.2	4.7	1.4	versicolor
64	3.2	4.5	1.5	versicolor
69	3.1	4.9	1.5	versicolor
62	2.8	4.8	1.8	virginica
61	3.0	4.9	1.8	virginica
64	2.8	5.6	2.1	virginica
72	3.0	5.8	1.6	virginica

1. (5 marks) He develops a nearest neighbour classification method. Given an unseen test instance, it calculates the Euclidean distance between the test instance and each of the 10 training instance, and set the predicted class label as the label of the training instance with the smallest distance. However, his classification accuracy was unsatisfactory. Find **two** possible reasons, and suggest a way to address each reason.

2. (5 marks) He develops a decision tree method (by discretising the feature values). The decision tree model has very high training accuracy. However, the test accuracy is poor. Find a possible reason, and suggest a way to address that.

Question 3. (10 marks)

Consider the following dataset describing 10 pizzas from a pizza shop, of which 5 are popular with customers, and 5 are not. They are described by three attributes: whether they have mushroom or not, whether they are vegetarian or not, and whether they are in small, medium or large size.

Instance	Mushroom	Vegetarian	Size	Class
1	yes	no	medium	popular
2	yes	no	large	popular
3	yes	no	medium	popular
4	yes	yes	large	popular
5	no	no	small	popular
6	no	no	large	unpopular
7	yes	yes	small	unpopular
8	no	yes	small	unpopular
9	no	no	medium	unpopular
10	no	yes	large	unpopular

The pizza shop wants to build a *decision tree* for classifying pizzas to *popular* or *unpopular*. Which attribute should be chosen for the *root* of the decision tree if they use the weighted average impurity function (the impurity measure is defined as $P(popular) * P(unpopular)$)? **Show your working.**

Part 4: Other Topics: Questions [15 marks]

1. (3 marks) Describe the main idea of Deep Learning. Give three examples of deep learning algorithms and their applications. Use no more than 100 words for this question.
2. (3 marks) Use no more than 150 words to describe the main ideas of Support Vector Machines for binary classification. Draw a figure to show your idea if necessary.
3. (3 marks) Text mining and natural language processing is a hot topic in Artificial Intelligence these days. Give two example applications/tasks in this topic, and state two good algorithms for tackling each of the application tasks. Use no more than 200 words for this question.
4. (3 marks) Knowledge-based systems are a fundamental topic of Artificial Intelligence. State three successful example systems and their applications that you have heard of/seen to date. Use no more than 100 words for this question.
5. (3 marks) Big data has been a very hot interdisciplinary topic recently. Use no more than 100 words to describe the five Vs of big data.

3 Notes

During the time between the assignment handout and submission, the tutor(s) will run a number of helpdesks to provide assistance.

4 Submission Guidelines

4.1 Submission Requirements

A document consisting of the answers of all the questions. The document can be written in PDF, text or the DOC format.

4.2 Submission Method

The PDF/Text/DOC version of the document should be submitted through web submission system from the COMP307/420 course web site **by the due time**.

4.3 Late Penalties

All assignments must be submitted on time unless you have made a prior arrangement with the lecturer or have a valid medical excuse (for minor illnesses it is sufficient to discuss this with the lecturer). The penalty for assignments that are handed in late without prior arrangement is one grade reduction per day. Assignments that are more than one week late will not be marked.