### ARTIFICIAL INTELIGENCE COMPSCI4004 2020/2021

### Author Guidelines for AI COMPSCI 4004 2020/2021 [INSERT YOUR OWN TITLE]

*Author(s) zhouyang zhen Name(s)zhouyang shen Student ID(s)2359009s*

University of Glasgow

#### Abstract

The abstract should appear at the top of the left-hand column of text, about 0.5 inch (12 mm) below the title area and no more than 3.125 inches (80 mm) in length. Leave a 0.5 inch (12 mm) space between the end of the abstract and the beginning of the main text. The abstract should contain about 100 to 150 words.

**Index Terms—** One, two, three, four, five

**1. Introduction**

These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts. Please follow them and if you have any questions, direct them to the course coordinator.

**2. Formatting your paper**

All printed material, including text, illustrations, and charts, must be kept within a print area of 7 inches (178 mm) wide by 9 inches (229 mm) high. Do not write or print anything outside the print area. The top margin must be 1 inch (25 mm), except for the title page, and the left margin must be 0.75 inch (19 mm). All *text* must be in a two-column format. Columns are to be 3.39 inches (86 mm) wide, with a 0.24 inch (6 mm) space between them. Text must be fully justified.

**3. Page title section**

The paper title (on the first page) should begin 1.38 inches (35 mm) from the top edge of the page, centered, completely capitalized, and in Times 14-point, boldface type. The authors’ name(s) and affiliation(s) appear below the title in ital and lower case letters.

**4. Type-style and fonts**

To achieve the best rendering, we strongly encourage you to use Times-Roman font. In addition, this will give the proceedings a more uniform look. Use a font that is no smaller than nine point type throughout the paper, including figure captions.

In nine point type font, capital letters are 2 mm high. If you use the smallest point size, there should be no more than 3.2 lines/cm (8 lines/inch) vertically. This is a minimum spacing; 2.75 lines/cm (7 lines/inch) will make the paper much more readable. Larger type sizes require correspondingly larger vertical spacing. Please do not double-space your paper. True-Type 1 fonts are preferred.

The first paragraph in each section should not be indented, but all following paragraphs within the section should be indented as these paragraphs demonstrate.

**5. MAjor headings**

Major headings, for example, “1. Introduction”, should appear in all capital letters, bold face if possible, centered in the column, with one blank line before, and one blank line after. Use a period (“.”) after the heading number, not a colon.

**5.1. Subheadings**

Subheadings should appear in lower case (initial word capitalized) in boldface.  They should start at the left margin on a separate line.

*5.1.1. Sub-subheadings*

Sub-subheadings, as in this paragraph, are discouraged. However, if you must use them, they should appear in lower case (initial word capitalized) and start at the left margin on a separate line, with paragraph text beginning on the following line.  They should be in italics.

6. Illustrations, graphs, and photographs

Illustrations must appear within the designated margins. They may span the two columns. If possible, position illustrations at the top of columns, rather than in the middle or at the bottom. Caption and number every illustration. All halftone illustrations must be clear black and white prints. Colors may be used, but they should be selected so as to be readable when printed on a black-only printer.

**7. Footnotes**

Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced. Use Times 9-point type, single-spaced. To help syour readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

**8. References**

List and number all bibliographical references at the end of the paper. The references can be numbered in alphabetic order or in order of appearance in the document. When referring to them in the text, type the corresponding reference number in square brackets as shown at the end of this sentence [1].

[1] A.B. Smith, C.D. Jones, and E.F. Roberts, “Article Title,” *Journal*, Publisher, Location, pp. 1-10, Date.

[2] Jones, C.D., A.B. Smith, and E.F. Roberts, *Book Title*, Publisher, Location, Date.

1. **Policy search with linear function approximation**

The performance of the policy search will also be compared with random agent and deterministic agent with baseline of fixed action 1, and in addition, for further evaluation, there will be performance

Comparison between existing agent and other agents, still, the metrics that has been performed for assessing are the total reward of the optimal policy, the analysis of the states graphs as well as the

average performance (e.g. reward) vs number of episodes.

* 1. **Comparison between random agent and deterministic agent**

First of all, we will compare the total reward of the agent with random agent and deterministic agent for each of the problems without noise, and then compare the agents given the problem and noisy environment, finally, the stochastic environment with and without noise will also be compared for evaluation.

After the training and prediction of the agent, the total reward of problem 0 without noise is calculated as

-1.284974326755993, whereas the total reward of random agent is -1.8417964935801103, the total reward of deterministic agent is -1. 6336942500717837.The results indicates that the policy gradient agent with function approximation has learned from previous 2000 episodes and performed the policy that has outweighed the performance of those 2 agents. Also, through observation from the actions that have been taken in episode, we discovered that the track and trace action could bring higher reward to problem 0 without noise applied. (graph)

For the problem 1 without noise, the total reward of policy search agent is computed as -0.39689804509900606, the reward of random agent is -0.6773462566122005, the reward of deterministic agent is still -1.6336942500717837,again,there is strong indication that the agent has learned well in training stage. For problem 2 without noise, the reward of policy search agent is -0.6293770338129265, the random agent has reward of -0.6181529019884758, the deterministic agent shares the reward of -1.6336934175695137 again by applying full lockdown every time. It seems that the random agent has exceeded the performance of policy search agent this time, Then, the state of reward graph is supplied to complementally evaluate how agent perform.

文本, 信件

描述已自动生成

文本, 信件, 白板

描述已自动生成

文本, 信件, 白板

描述已自动生成

If we observe the tables above, we could find that all 3 agents has achieved decent states, However, the policy agent has gradually taken the higher reward with weeks go by, whereas the random agent has randomly selected the actions to improve the total reward, then, to understand whether the total reward of policy search is

reasonable, the learning converging graph is also provided:

图表, 折线图

描述已自动生成

As the figure showing above, we can see that the converging has not occurred properly, since the reward difference between random agent and policy search agent is not massive, we could therefore deduce that the agent has not learned useful information at this problem, the reason could be the improper use of learning rate..

For the problem 3 without noise, the policy gradient agent has reached the total reward of -0.6317200225475976, total reward of

-0.7683753703346045 is performed by random agent and that of -1.6336934175695137 is still done by deterministic agent,

For problem 4 without noise, the policy search has gained -1.459818869796979, In contrast, the random agent has come done with the reward of -1.2250992288646896, reward of

-1.63369396954822 is operated by deterministic agent. Since the reward of policy search is not optimistic, the analysis of the state and reward graph needed to be executed,

文本, 信件, 白板

描述已自动生成

图片包含 图表

描述已自动生成

文本, 信件, 白板

描述已自动生成

As we can observe from the tables above, the policy search agent has reduced the infected and quarantined population to the minimal ,the reason is that it executed full lockdown at almost each step, which could be observed from figure below showing the policy executed:

Moreover, through observing the converging graph of this problem. Which is displayed below, it is clear to say that the agent has learn well

图片包含 图表

描述已自动生成

For problem 5 without the noise, the current agent has achieved total reward of

-1.4358062007251744, whereas the random and deterministic agent have operated total reward of -1.605427938929334 and -1.6336946050766263.

For problem 6 without noise, the policy gradient has found the optimal policy sharing the same actions performed by deterministic, which is -1.633694892623494,the random agent however has obtained reward of -1.9217016717205564,it indicate that for this specific problem and environment, the policy gradient agent has discovered that the whole episode lockdown could help with increasing reward.

For problem 7 without noise, total reward of -1.5712610153381423 is achieved by policy gradient agent, total reward of -2.548581789279756 and -1.6336951833767266 is performed by random agent and deterministic agent.

For problem 8 without noise, deterministic agent has achieved the better performance with total reward of

-1.6336954773902277, whereas policy search and random agent has only reached -1.9855363835376723 and

-2.1736730905057637, again, for further analysis, the state and reward graph, actions and learning graph for deterministic and policy search agent is presented.

图片包含 文本

描述已自动生成

文本, 信件, 白板

描述已自动生成

Action graph for ps\_lfa

图片包含 图表

描述已自动生成As we can observe from the graphs, the deterministic agent achieved the better performance and controlled the quarantined and infected person by always applying full lockdown at each step, where as the policy search agent leaned from episodes and think that by applying track and trace at each step of episode might be the optimal choice for this specific problem.