基于强化学习和马尔可夫模型的出租车代理

Xiuyuan Qi
qixy1@shanghaitech.edu.cn

Ziyang Guo
guozy@shanghaitech.edu.cn

Abstract

本项目为使用人工智能完成一个出租车游戏，着重研究了三种出租车代理程序：Search Agent, Reinforcement Agent 和 Markov-Search Agent, 并对它们的不同表现进行比较。 Search Agent采用广度优先搜索算法；Reinforcement Agent采用强化学习算法；Markov-Search Agent 使用隐马尔科夫模型进行决策。 结果说明，Search Agent 一定得到最优解；（你来写）；而在加入了迷雾和天气系统的游戏中，Markov-Search Agent 有着不错的表现。

1 问题描述

在一个$M \times N$ 的地图中，每个格子与相邻的四个格子连通，如果相邻的两个格子之间存在墙壁，则两者不再连通。

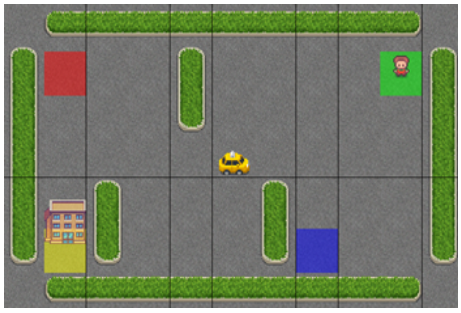


Figure 1: A possible map.

每一轮有1名乘客和1个乘客想前往的目的地。

地图中一共有4个乘车点，每局游戏开始时乘客和目的地会随机刷新在不同的乘车点。出租车会随机出生在地图中。

出租车每步可以进行上/下/左/右移动（从一个格子移动到与之连通的另一个格子），或者进行载客/下客操作。

当出租车位于乘客所在的乘车点，且乘客不在车内时，进行上车操作会使乘客转移到车中。否则操作无效。

当出租车位于目的地，且乘客在车内时，进行下车操作会使乘客转移到车所在的地块，并且本局游戏结束。否则操作无效。

得分规则

- 无效的上下车：-10 （与乘客不在同一个格子的情况下上车、车上无乘客或不在目的地时下车）
- 将乘客送达目的地：+20（乘客在车中、车在目的地时进行下车操作）

Preprint. Under review.

- 其他：-1 （移动、合法上下车）

Agent 的目的是使得分最大化，即以尽量少的步数将乘客送达目的地。

若出租车未能在200步内将乘客送达目的地，则本局游戏将强制结束，以最终得分为本局得分。

1.1 附加规则

为了增加难度与不确定性、以及添加前后局之间的关联性，对于部分游戏局，我们添加了以下的附加规则：

1.1.1 迷雾

在添加了迷雾的游戏中，存在一个额外的参数V ，只有当乘客与出租车的横、纵距离均小于V时，出租车才能收到乘客的位置信息，否则出租车无法知道乘客的位置。

1.1.2 天气

在连续的多局游戏中，存在一个参数“天气”W。 共有3种天气：晴天($W=0$)、阴天($W=1$)、雨天($W=2$)。

每局天气固定。 本局的天气决定下一局各天气出现的概率， 同时决定本局中乘客在各乘车点的出现概率。

暂定天气转移矩阵T如下：

$$T = P(W_{t+1}|W_t) = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.15 & 0.4 & 0.45 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}$$

其中W_t为第t局时的天气变量。设W_t的取值为w_t，则$(T)_{w_t,w_{t+1}} = P(w_{t+1}|w_t)$，即$T$中第$w_t$行第$w_{t+1}$列元素为本局天气为$w_t$的情况下下一局天气为$w_{t+1}$的概率。

暂定天气影响乘客概率的矩阵如下：

$$P(L_t|W_t) = \begin{pmatrix} 0.2 & 0.1 & 0.1 & 0.6 \\ 0.1 & 0.4 & 0.4 & 0.1 \\ 0.7 & 0.1 & 0.1 & 0.1 \end{pmatrix}$$

其中L_t为第t局中的乘客所在乘车点变量，即乘客出现在了4个乘车点中的哪一个。

暂定初始天气分布$P(W_0)$如下：

$$P(W_0) = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

即第0局中3种天气的出现概率均等。

2 使用的模型

本项目着重研究了三种游戏 Agent 模型：Search Agent, Reinforcement Agent 和 Markov-Search Agent.

2.1 Search Agent

Search Agent 在拥有完整信息的地图中活动（Agent 得知乘客与目的地的确切位置）。

Search Agent 采用广度优先搜索算法 (BFS)，先搜索一条通往乘客位置的路径，沿路径前往上车点进行上车操作后，再搜索一条通往目的地的路径，前往目的地让乘客下车。

2.2 Reinforcement Agent

同 Search Agent 一样， Reinforcement Agent 也得知乘客与目的地位置。

Reinforcement Agent 采用了Q-Learning算法，其中与课堂中例子的区别是存在多个结束状态， 算法上采用observation state-action pair作为Q table的索引。相比可以画在地图上的Q Table ，这样会形成一个多维的Q Table，更类似吃豆人的例子。

其次，采用exploration function的方式来鼓励探索。更新Q value的公式如下:

$$Q(s,a) = \alpha R(s,a,s') + \gamma max_{a'} Q(s',a') + \frac{explore\_constant}{N(s,a)}$$

其中， $N(s,a)$ 对应在 s状态下选择a行动的次数；explore_constant 为超参数，默认为1。

另外，我们在编写对应的函数时还考虑了训练时间的问题：为了避免在学习次数或其他参数改变 时需要较多时间重新训练模型，函数中支持传入已经

2.3 Markov-Search Agent

在附加规则下，天气系统可看作是一个隐马尔科夫模型 (HMM)，能以贝叶斯网络的形式表示为下图:

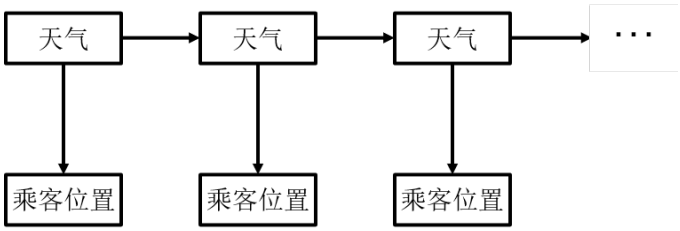


Figure 2: Weather HMM

Markov-Search Agent 可基于已知信息，计算出当前游戏局乘客在各乘车点的出现概率，据此决定出租车的行动。具体算法如下:

1. 使用 Filtering Algorithm 计算当前各天气的概率:

$$P(W_t|l_{0:t}) = \alpha P(l_t|W_t) * (P(W_{t-1}|l_{0:t-1})P(W_t|W_{t-1}))$$

   其中$*$号表示矩阵对应元素相乘，不显示符号的乘法表示矩阵乘法。α为归一化系数，使得结果向量中的各元素和为1.
   l_t表示L_t的取值。$l_{0:t}$等同于$l_0, l_1, \ldots, l_t$.
2. 计算当前乘客出现在各乘车点的概率:

$$P(L_t|l_{0:t-1}) = P(W_{t-1}|l_{0:t-1})P(W_t|W_{t-1})P(L_t|W_t)$$

3. 根据概率计算每个位置的期望得分
4. 搜索一条通往期望最高的乘车点的路径；如果到了附近看到没有乘客，则搜索通往期望第二高的乘车点的路径，如此以往

3 表现评估

3.1 Search Agent

由于 BFS 算法性质保证了搜索结果为最短路径，且载客、送客两步均不可跳过，可知 Search Agent 必然能得到最优解。因此 Search Agent 可作为另外两种 Agent 的表现参考。

3.2 Reinforcement Agent

3.2.1 学习过程

在本游戏中，如果agent成功将乘客送达目的地，其得分一般为正数。 并且由于存在200次的时限，agent的得分区间为[-2000,20)。在训练过程中可以发现，一个常见的失败情况是

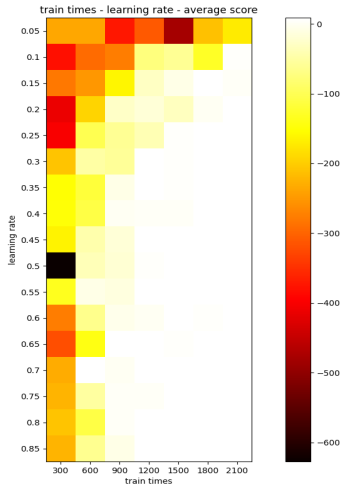


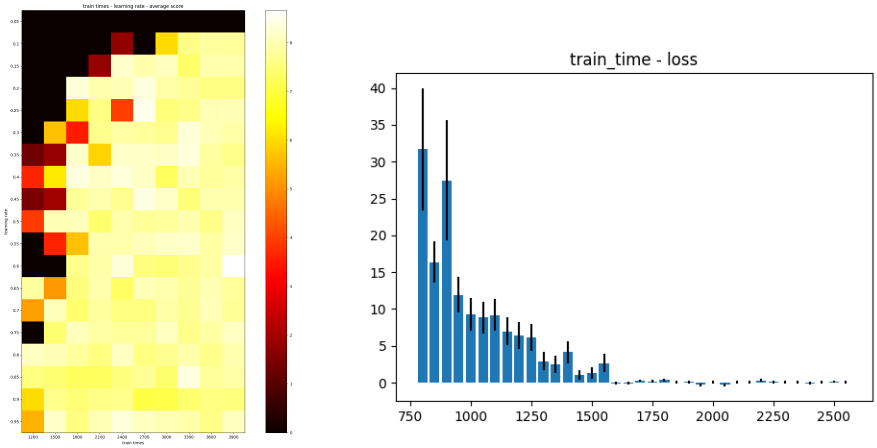
Figure 3: scores under different learning rate and training times.

agent一直向着某个方向移动，并在-200分时退出本局。将不同学习率下agent训练不同次数的得分情况 可视化后得到下图。

从图中可以看出：随着学习率下降，达到正得分并收敛所需的训练次数逐渐减少。 然而由于负数部分绝对值较大，较难判断完成游戏时采取的策略的好坏（即正分数的相对大小）因此采用

$$f(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$$

函数来对得分进行处理，结果如下图左。收敛后的得分基本在8分左右，观察学习率为0.4的损失曲线 得知在约1600次训练后loss基本稳定为零，即收敛到了最优解。



(a) scores under relu function.

(b) loss figure with error

Figure 4: scores and loss

3.2.2 表现对比

在本游戏中，由于Search Agent一定会得到最优解，因此损失曲线同时也是其他Agent与Search Agent的得分差距。学习率>0.2 时Reinforcement Agent 基本都可以在2000次迭代后收敛 并达到与Search Agent相同的最优效果。

另外，在未收敛的情况下，很大一部分得分为-200；也有少部分情况有更低分。随着训练次数增加 得分一般会呈现先快后慢的增长趋势。

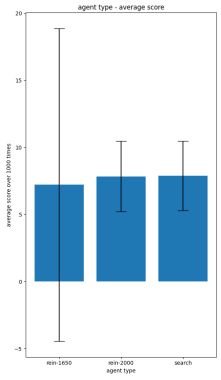


Figure 5: scores under different learning rate and training times.

此外，我们还对游戏环境进行了一定的改动，来测试在更大的地图下该AI的表现。 大地图的字符串如下

```
"+-----------------------+",
"|␣|R:␣:␣:␣:␣|␣:␣:␣|␣:␣:␣|",
"|␣|␣:␣|␣:␣:␣|␣:␣:G|␣:␣:␣|",
"|␣:␣:␣|␣|␣:␣:␣:␣:␣|␣:␣:␣|",
"|␣:␣:␣|␣|␣|␣:␣:␣:␣|␣:␣:␣|",
"|␣:␣:␣:␣|␣|␣|␣:␣:␣|B|␣:␣|",
"|␣:␣:␣:␣|␣:␣|␣:␣:␣:␣|␣|␣|",
"|␣:␣:␣:Y|␣:␣|␣:␣:␣:␣:␣:␣|",
"+-----------------------+",
```

其中，| 代表墙壁; : 代表可以通过; R,G,B,Y代表四个上下车点。在经过测试后发现 其学习过程的得分曲线与小地图中基本一致：都是类似反比函数的形式； 且随学习率提高收敛所需训练次数逐渐降低，并在0.5左右达到最低值。

3.3 **Markov-Search Agent** 在附加规则下的表现

4 主要贡献

郭子杨：Search Agent；天气系统；Markov-Search Agent

齐修远：基础环境配置；迷雾系统；强化学习Agent；表现可视化

项目GitHub仓库：`https://github.com/PandragonXIII/AI-team-project`

5 参考文献

Gymnasium API: `https://gymnasium.farama.org/environments/toy_text/taxi/`

6 Submission of papers to NeurIPS 2023

Please read the instructions below carefully and follow them faithfully. **Important:** This year the checklist will be submitted separately from the main paper in OpenReview, please review it well ahead of the submission deadline: `https://neurips.cc/public/guides/PaperChecklist`.

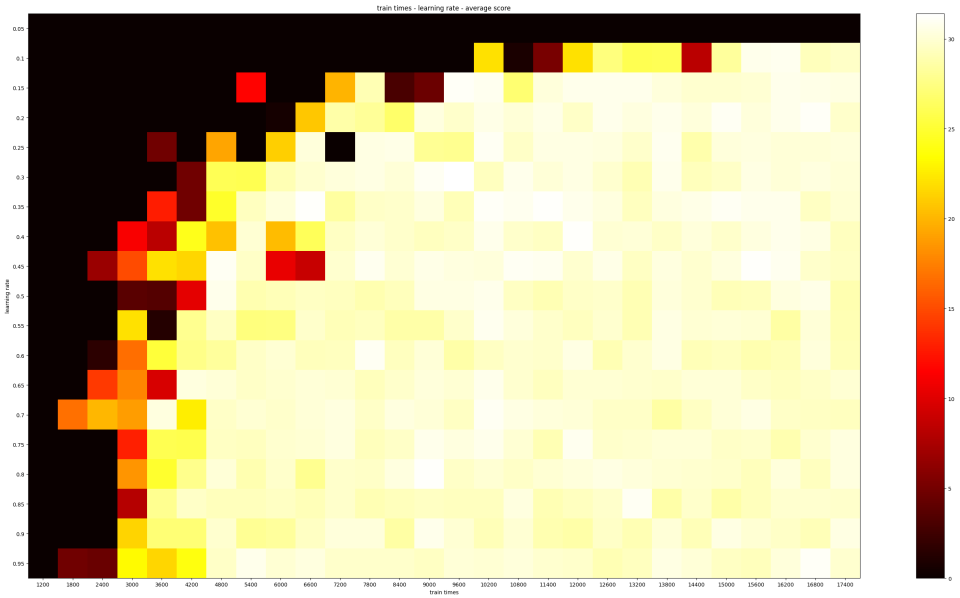


Figure 6: scores on big map after ReLU.

6.1 Style

Papers to be submitted to NeurIPS 2023 must be prepared according to the instructions presented here. Papers may only be up to **nine** pages long, including figures. Additional pages *containing only acknowledgments and references* are allowed. Papers that exceed the page limit will not be reviewed, or in any other way considered for presentation at the conference.

The margins in 2023 are the same as those in previous years.

Authors are required to use the NeurIPS LaTeX style files obtainable at the NeurIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

6.2 Retrieval of style files

The style files for NeurIPS and other conference information are available on the website at

`http://www.neurips.cc/`

The file `neurips_2023.pdf` contains these instructions and illustrates the various formatting requirements your NeurIPS paper must satisfy.

The only supported style file for NeurIPS 2023 is `neurips_2023.sty`, rewritten for LaTeX 2_ε. **Previous style files for LaTeX 2.09, Microsoft Word, and RTF are no longer supported!**

The LaTeX style file contains three optional arguments: `final`, which creates a camera-ready copy, `preprint`, which creates a preprint for submission to, e.g., arXiv, and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

Preprint option If you wish to post a preprint of your work online, e.g., on arXiv, using the NeurIPS style, please use the `preprint` option. This will create a nonanonymized version of your work with the text "Preprint. Work in progress." in the footer. This version may be distributed as you see fit, as long as you do not say which conference it was submitted to. Please **do not** use the `final` option, which should **only** be used for papers accepted to NeurIPS.

At submission time, please omit the `final` and `preprint` options. This will anonymize your submission and add line numbers to aid review. Please do *not* refer to these line numbers in your paper as they will be removed during generation of camera-ready copies.

The file `neurips_2023.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in Sections 7, 8, and 9 below.

7 General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by ½ line space (5.5 points), with no indentation.

The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow ¼ inch space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors' names are set in boldface, and each name is centered above the corresponding address. The lead author's name is to be listed first (left-most), and the co-authors' names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in Section 9 regarding figures, tables, acknowledgments, and references.

8 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

8.1 Headings: second level

Second-level headings should be in 10-point type.

8.1.1 Headings: third level

Third-level headings should be in 10-point type.

Paragraphs There is also a `\paragraph` command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

9 Citations, figures, tables, references

These instructions apply to everyone.

9.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

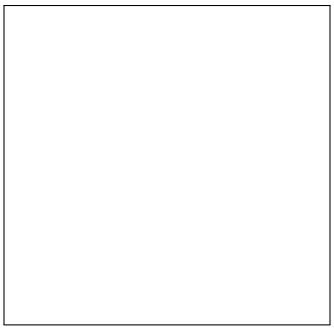


Figure 7: Sample figure caption.

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

> Hasselmo, et al. (1995) investigated…

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2023` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2023}
```

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]," not "In our previous work [4]." If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form "A. Anonymous" and include a copy of the anonymized paper in the supplementary material.

9.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.[2]

9.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

[1] Sample of the first footnote.

[2] As in this example.

Table 1: Sample table title

	Part	
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

9.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

`https://www.ctan.org/pkg/booktabs`

This package was used to typeset Table 1.

9.5 Math

Note that display math in bare TeX commands will not create correct line numbers for submission. Please use LaTeX (or AMSTeX) commands for unnumbered display math. (You really shouldn't be using $$ anyway; see `https://tex.stackexchange.com/questions/503/why-is-preferable-to` and `https://tex.stackexchange.com/questions/40492/what-are-the-differences-between-align-equation-and-displaymath` for more information.)

9.6 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

10 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4."

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

  ```
  \usepackage{amsfonts}
  ```

  followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

```
\newcommand{\RR}{I\!\!R} %real numbers
\newcommand{\Nat}{I\!\!N} %natural numbers
\newcommand{\CC}{I\!\!\!\!C} %complex numbers
```

Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

10.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (`http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

Acknowledgments and Disclosure of Funding

Use unnumbered first level headings for the acknowledgments. All acknowledgments go at the end of the paper before the list of references. Moreover, you are required to declare funding (financial activities supporting the submitted work) and competing interests (related financial activities outside the submitted work). More information about this disclosure can be found at: `https://neurips.cc/Conferences/2023/PaperInformation/FundingDisclosure`.

Do **not** include this section in the anonymized submission, only in the final paper. You can use the `ack` environment provided in the style file to autmoatically hide this section in the anonymized submission.

11 Supplementary Material

Authors may wish to optionally include extra information (complete proofs, additional experiments and plots) in the appendix. All such materials should be part of the supplemental material (submitted separately) and should NOT be included in the main submission.

References

References follow the acknowledgments in the camera-ready paper. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. Note that the Reference section does not count towards the page limit.

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.