
Hierarchical Reinforcement Learning using Auxiliary Reward

Paper ID 812

Abstract

The abstract paragraph should be indented 1/2 inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Method

1.1 Set Auxiliary Reward

We use the call-and-return model to execute skills. Once selected, the skill is executed for k steps. The TD residual of high-level value function V_{π_h} with discount γ is defined as $\delta_t^V = r_t^h + \gamma V_{\pi_h}(s_{t+k}^h) - V_{\pi_h}(s_t^h)$ [1], where r_t^h is the cumulative k -step reward, $r_t^h = \sum_{i=0}^{k-1} r_{t+i}$. The high-level TD residual can be considered as an estimation of the advantage of the action a_t^h .

$$\begin{aligned} A_{\pi_h}(s_t^h, a_t^h) &= \mathbb{E}_{s_{t+k}^h} [Q_{\pi_h}(s_t^h, a_t^h) - V_{\pi_h}(s_t^h)] \\ &= \mathbb{E}_{s_{t+k}^h} [r_t^h + \gamma V_{\pi_h}(s_{t+k}^h) - V_{\pi_h}(s_t^h)]. \end{aligned}$$

When the environment reward is sparse or irrelevant to time-step t , r_t^h can be omitted from the advantage function, and $A_{\pi_h}(s_t^h, a_t^h)$ can be estimated as $\mathbb{E}_{s_{t+k}^h} [\gamma V_{\pi_h}(s_{t+k}^h) - V_{\pi_h}(s_t^h)]$. To guarantee that the auxiliary reward can facilitate the learning of the high-level policy π_h , we set the estimated high-level advantage function as our auxiliary reward to the low-level policy.

$$R_l(s_t^l, s_{t+k-1}^l) = \gamma V_{\pi_h}(s_{t+k}^h) - V_{\pi_h}(s_t^h),$$

where $R_l(s_t^l, s_{t+k-1}^l)$ denotes the total k -step auxiliary reward. As each low-level skill produces a consistent effect in the task, we split the total auxiliary reward equally to each low-level step.

$$R_l(s_t^l) = \frac{\gamma V_{\pi_h}(s_{t+k}^h) - V_{\pi_h}(s_t^h)}{k}. \quad (1)$$

When optimizing the low level policy using policy gradient method, we maximize the performance measure $\mathbb{E}_{(s_t^h, s_t^l, a_t^h, a_t^l) \sim (\pi_h, \pi_l)} [R_l(s_t^l, a_t^l)]$. Based on Equation 1,

$$\mathbb{E}_{(s_t^h, s_t^l, a_t^h, a_t^l) \sim (\pi_h, \pi_l)} [R_l(s_t^l, a_t^l)] = \frac{1}{k} \mathbb{E}_{(s_t^h, s_t^l, a_t^h, a_t^l) \sim (\pi_h, \pi_l)} [A_{\pi_h}(s_t^h, a_t^h)] \quad (2)$$

$$= \frac{1}{k} \mathbb{E}_{(s_t^h, a_t^h) \sim \pi_h} [\mathbb{E}_{(s_t^l, a_t^l | s_t^h, a_t^h) \sim \pi_l} [A_{\pi_h}(s_t^h, a_t^h)]] \quad (3)$$

$$= \frac{1}{k} \mathbb{E}_{(s_t^h, a_t^h) \sim \pi_h} [A_{\pi_h}(s_t^h, a_t^h)] \cdot \mathbb{E}_{(s_t^l, a_t^l | s_t^h, a_t^h) \sim \pi_l} [1] \quad (4)$$

$$= \frac{1}{k} \mathbb{E}_{(s_t^h, a_t^h) \sim \pi_h} [A_{\pi_h}(s_t^h, a_t^h)]. \quad (5)$$

So optimizing the performance measure of the low-level auxiliary reward is equal to optimizing the performance measure of the environment reward. With fixed π_h , the objective of learning π_l with our auxiliary reward R_l is maximizing the the return of the joint policy π of π_h and π_l . We do a one-step estimation of the advantage function in Equation 2.

23 1.2 Style

24 Papers to be submitted to NeurIPS 2019 must be prepared according to the instructions presented
25 here. Papers may only be up to eight pages long, including figures. Additional pages *containing only*
26 *acknowledgments and/or cited references* are allowed. Papers that exceed eight pages of content
27 (ignoring references) will not be reviewed, or in any other way considered for presentation at the
28 conference.

29 The margins in 2019 are the same as since 2007, which allow for $\sim 15\%$ more words in the paper
30 compared to earlier years.

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

34 1.3 Retrieval of style files

35 The style files for NeurIPS and other conference information are available on the World Wide Web at

36 <http://www.neurips.cc/>

37 The file `neurips_2019.pdf` contains these instructions and illustrates the various formatting re-
38 quirements your NeurIPS paper must satisfy.

39 The only supported style file for NeurIPS 2019 is `neurips_2019.sty`, rewritten for L^AT_EX 2_ε.
40 **Previous style files for L^AT_EX 2.09, Microsoft Word, and RTF are no longer supported!**

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

44 **Preprint option** If you wish to post a preprint of your work online, e.g., on arXiv, using the
45 NeurIPS style, please use the `preprint` option. This will create a nonanonymized version of your
46 work with the text “Preprint. Work in progress.” in the footer. This version may be distributed as
47 you see fit. Please **do not** use the `final` option, which should **only** be used for papers accepted to
48 NeurIPS.

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

52 The file `neurips_2019.tex` may be used as a “shell” for writing your paper. All you have to do is
53 replace the author, title, abstract, and text of the paper with your own.

54 The formatting instructions contained in these style files are summarized in Sections 2, 3, and 4
55 below.

56 2 General formatting instructions

57 The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long.
58 The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points.
59 Times New Roman is the preferred typeface throughout, and will be selected for you by default.
60 Paragraphs are separated by $\frac{1}{2}$ line space (5.5 points), with no indentation.

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

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

69 Please pay special attention to the instructions in Section 4 regarding figures, tables, acknowledgments,
70 and references.

71 **3 Headings: first level**

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

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

74 **3.1 Headings: second level**

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

76 **3.1.1 Headings: third level**

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

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

80 **4 Citations, figures, tables, references**

81 These instructions apply to everyone.

82 **4.1 Citations within the text**

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

86 The documentation for `natbib` may be found at

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

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

90 `\citet{hasselmo}` investigated\dots

91 produces

92 Hasselmo, et al. (1995) investigated...

93 If you wish to load the `natbib` package with options, you may add the following before loading the
94 `neurips_2019` package:

95 `\PassOptionsToPackage{options}{natbib}`

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

98 `\usepackage[nonatbib]{neurips_2019}`

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

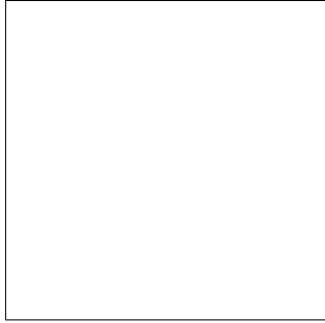


Figure 1: Sample figure caption.

Table 1: Sample table title

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

103 4.2 Footnotes

104 Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number¹
 105 in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote
 106 with a horizontal rule of 2 inches (12 picas).

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

108 4.3 Figures

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

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

115 4.4 Tables

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

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

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

123 <https://www.ctan.org/pkg/booktabs>

124 This package was used to typeset Table 1.

¹Sample of the first footnote.

²As in this example.

5 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.

6 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 file uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdf fonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NeurIPS. Please see <http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf>
- `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}{\mathbb{R}} %real numbers
\newcommand{\Nat}{\mathbb{N}} %natural numbers
\newcommand{\CC}{\mathbb{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.

6.1 Margins in L^AT_EX

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 L^AT_EX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References

- [1] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.