

Accelerating Entropy-Based Transformer Calibration

Michael Hu

Adviser: Prof. Karthik Narasimhan

Abstract

This independent work tackles the problem of entropy rate drift in open-ended text generation. Entropy rate drift is a phenomenon where the entropy rate of a language model deviates from the entropy rate of the context seed over time. Here, we give empirical evidence that entropy rate drift is a consequence of sampling from the tail of a probability distribution. Next, we show that the lookahead entropies of tokens in the tail are markedly higher than the most common k . We use these results to motivate the usage of one-step lookahead calibration, and show ways to make the technique faster and more effective. Finally, we demonstrate ways that the technique can be connected to other heuristics that improve open-ended text generation.

1. Introduction

This document is a sample you can use for formatting your independent work report. An example template with all bib and supporting files can be downloaded from the IW website at <http://iw.cs.princeton.edu>.

2. Preparation Instructions

2.1. Paper Formatting

There are no minimum or maximum length limits on IW reports. We are including this template because we think it will be helpful for citing things properly and for including figures into formatted text. If you are using L^AT_EX [1] to typeset your paper, then we strongly suggest that you start from the template available at <http://iw.cs.princeton.edu> – this document was prepared with that template. If you are using a different software package to typeset your paper, then you can still use this

document as a reasonable sample of how your report might look. Table 1 is a suggestion of some formatting guidelines, as well as being an example of how to include a table in a Latex document.

Field	Value
Paper size	US Letter 8.5in \times 11in
Top margin	1in
Bottom margin	1in
Left margin	1in
Right margin	1in
Body font	12pt
Abstract font	12pt, italicized
Section heading font	14pt, bold
Subsection heading font	12pt, bold

Table 1: Formatting guidelines.

Please ensure that you include page numbers with your submission. This makes it easier for readers to refer to different parts of your paper when they provide comments.

We highly recommend you use bibtex for managing your references and citations. You can add bib entries to a references.bib file throughout the semester (e.g., as you read papers) and then they will be ready for you to cite when you start writing the report. If you use bibtex, please note that the references.bib file provided in the template example includes some format-specific incantations at the top of the file. If you substitute your own bib file, you will probably want to include these incantations at the top of it.

2.2. Citations and Footnotes

There are various reasons to cite prior work and include it as references in your bibliography. For example, If you are improving upon prior work, you should include a full citation for the work in the bibliography [2, 3]. You can also cite information that is used as background or explanation[4]. In addition to citing scholarly papers or books, you can also create bibtex entries for webpages or other sources. Many online databases allow you to download a premade bibtex entry for each paper you access. You can simply copy-paste these into your references.bib file.

Sometimes you want to footnote something, such as a web site.¹ Note that the footnote number comes after the punctuation.

2.3. Figures and Tables.

Figure 1 shows an example of how to include a figure in your report. Ensure that the figures and tables are legible. Please also ensure that you refer to your figures in the main text. Make sure that your figures will be legible in the expected forms that the report will be read. If you expect someone to print it out in gray-scale, then make sure the figures are legible when printed that way.

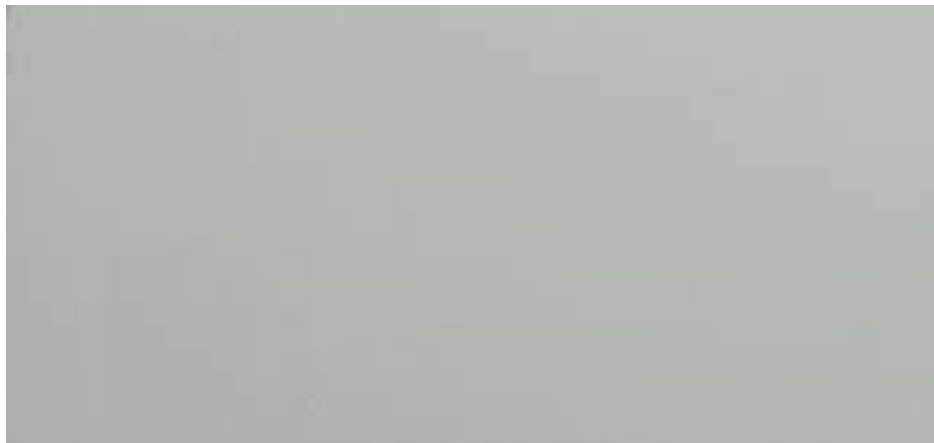


Figure 1: This is a gray image.

In Section 2.1, an example of a table was given. (Note that the “S” in Section is capitalized. Here’s one more example - see Table 2.

Some field	Another field
200	10000
400	20000
800	40000
1600	80000
3200	160000
6400	320000

Table 2: Some data in a table.

Here’s an example that shows how you can have side-by-side figures - see Figure 2 and Figure 3. (Note that the the “F” in Figure is capitalized.

¹<http://www.cs.princeton.edu>

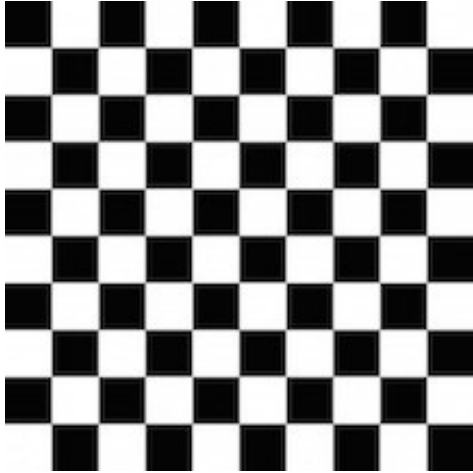


Figure 2: Plain checkerboard.

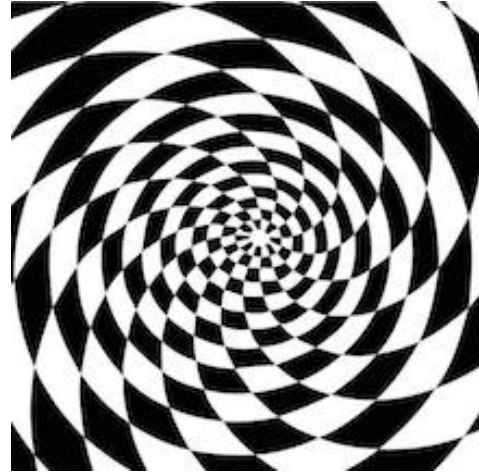


Figure 3: Cool checkerboard.

2.4. Double Quotes.

Latex double quotes are not the same as the double quote key on your keyboard. The standard way of writing quotes and double quotes in LaTeX is with “ and ” not with " and ".

Now that may be confusing, so you may want to use the `\{quotes}` command. For example “The quick brown fox.”

2.5. Main Body.

Avoid bad page or column breaks in your main text, i.e., last line of a paragraph at the top of a column or first line of a paragraph at the end of a column. If you begin a new section or sub-section near the end of a column, ensure that you have at least two (2) lines of body text on the same column.

3. Outline

The following is a possible outline for your paper.

3.1. Introduction

- Motivation and Goal (The goal of this project is...)
- Overview of challenge and previous work
- Approach

- Summary of implementation
- Summary of results

3.2. Problem Background and Related Work

- Survey of prior work with similar goals
- For each previous approach, explain what has been done and why it does not meet your goal

3.3. Approach

- Key novel idea
- Why it is a good idea

3.4. Implementation

- System overview (flow chart of key steps?)
- Subsection for each step or issue you addressed
 - Problem statement
 - Possible approaches
 - Chosen approach and why
 - Implementaton details

3.5. Evaluation

- Experiment design...
- Data...
- Metrics...
- Comparisons...
- Qualitative results...
- Quantitative results...

3.6. Summary

- Conclusions...

- Limitations...
- Future work...

4. Ethics

Your independent work report should abide by the basic standards of scholarly ethics and by the Princeton Honor Code. If you have any doubts about how to cite other work, how to quote or include text or images from other works, or other issues, please discuss them with your project adviser or with the IW coordinators.

References

- [1] L. Lamport, *LaTeX: A Document Preparation System*, 2nd ed. Reading, Massachusetts: Addison-Wesley, 1994.
- [2] F. Lastname1 and F. Lastname2, “A very nice paper to cite,” in *International Symposium on Computer Architecture*, 2000.
- [3] F. Lastname1 and F. Lastname2, “Another very nice paper to cite,” in *International Symposium on Computer Architecture*, 2001.
- [4] B. Salzberg and T. Murphy, “Latex: When Word fails you,” in *Proceedings of the 33rd Annual ACM SIGUCCS Conference on User Services*, ser. SIGUCCS ’05. New York, NY, USA: ACM, 2005, pp. 241–243.