Concordant Decoding

Gregory Lim and Second Author and Third Author
Singapore University of Technology and Design
{gregory_lim, ...}@mymail.sutd.edu.sg

Abstract

Main contributions and findings (200 words).

1 Introduction

Blah.

2 Related Work

Blah.

3 Methodology

Let V be a finite vocabulary totally ordered by $\preceq$, and $\Pr$ be a next-token kernel of an autoregressive language model. Let ℓ denote $\log \Pr$, and juxtaposition denote concatenation.

We define the decoding rule by

$$\arg\max_{y \in V} \left[\ell(y \mid x) + D(p \,\|\, q)\right] \tag{1}$$

where D is KL-divergence, and

$$p = \{\Pr(z \mid xy)\}_{z \in V},$$

and

$$q = \{\Pr(z \mid x)\}_{z \in V}.$$

4 Experimentation

5 Conclusion

References

John Doe and Jane Roe. 2025. An example paper. *Journal of Examples*.

A Appendix Title

Appendix content goes here (Doe and Roe, 2025).