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, and $\Delta : \mathcal{P}(V) \times \mathcal{P}(V) \to [0, \infty]$ be the Kullback–Leibler (KL) divergence function. Let ℓ denote $\log \Pr$, and juxtaposition denote concatenation.

We define the decoding rule by

$$f(x) \in \arg\max_{y \in V} \left[\ell(y \mid x) + \Delta(\hat{xy} \parallel \hat{x})\right] \qquad (1)$$

where $\hat{p}$ is the probability mass function satisfying

$$\hat{p}(v) = \Pr(v \mid p)$$

for each $v \in V$.

Fix a uniform-cost random-access machine model of computation.

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).