Amortizing Intractable Inference in Large Language Models

Edward J. Hu[*], Moksh Jain[*], Eric Elmoznino
Mila – Quebec AI Institute, Université de Montréal
{edward.hu,moksh.jain,eric.elmoznino,...

Younesse Kaddar[∞]
University of Oxford
younesse.kaddar@chch.ox.ac.uk

Guillaume Lajoie[†], Yoshua Bengio[◇], Nikolay Malkin
Mila – Quebec AI Institute, Université de Montréal
...,guillaume.lajoie,yoshua.bengio,nikolay.malkin}@mila.quebec

Posterior

翻译任务

$$q(Z \mid X) \propto p_{\mathrm{LM}}(XZ)^{1/T}$$

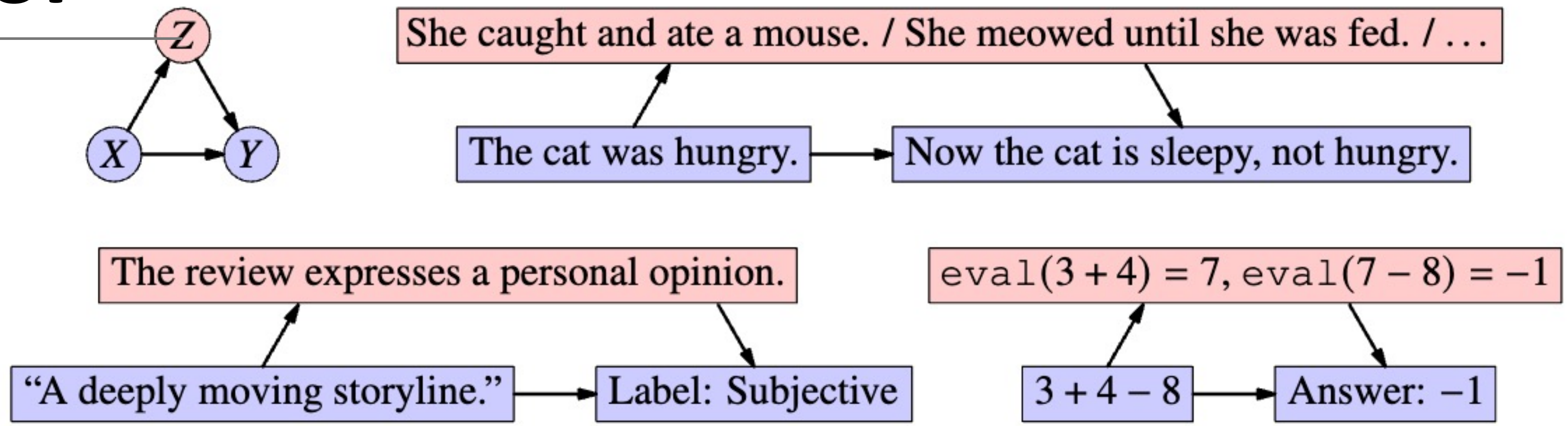
特有段落生成

$$q(Z \mid X) \propto p_{\mathrm{LM}}(XZ)^{\alpha} p_{\mathrm{LM}}(Z)^{\beta} \text{ with } \beta < 0 \text{ and } \alpha > 0,$$

受限生成

$$q(Z) \propto p_{\mathrm{LM}}(Z) c(Z)$$

Posterior

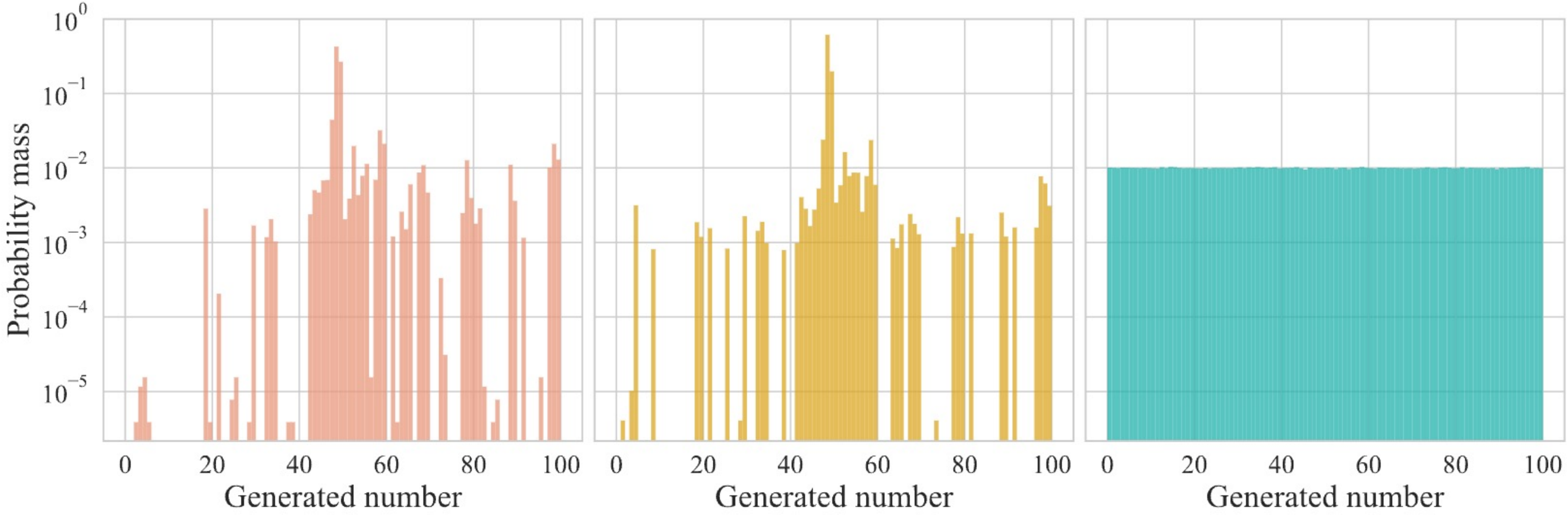


思维链 $q(Z \mid X, Y) \propto p_{\mathrm{LM}}(XZY)$

Object	Meaning	Example 1 (infilling)	Example 2 (subjectivity classification)
X	cause / condition / question	*The cat was hungry.*	*A deeply moving storyline.*
Z	mechanism / reasoning chain	*She ate a mouse.*	*This review expresses personal feelings.*
Y	effect / answer	*Now the cat is sleepy, not hungry.*	*Answer: Subjective*
$p(Z \mid X)$	conditional prior	$p_{\mathrm{LM}}(Z \mid X)$	
$p(Y \mid X, Z)$	likelihood of effect given cause and mechanism	$p_{\mathrm{LM}}(Y \mid XZ)$	
$p(Z, Y \mid X)$	conditional joint, reward for Z	$p_{\mathrm{LM}}(ZY \mid X)$	
$p(Z \mid X, Y)$	posterior **(intractable!)**	approximated and amortized by GFlowNet $q_{\mathrm{GFN}}(Z \mid X[, Y])$	
$q(Y \mid X)$	posterior predictive / Bayesian model average	approximated as $\sum_Z q_{\mathrm{GFN}}(Z \mid X) p_{\mathrm{LM}}(Y \mid XZ)$, sampled as $Z \sim q_{\mathrm{GFN}}(Z \mid X), Y \sim p_{\mathrm{LM}}(Y \mid XZ)$	

Model

Prompt: The following is a random integer drawn uniformly between 0 and 100



(a) **Base model**
50.5% of samples
are valid numbers.

(b) **PPO fine-tuning**
95.8% of samples
are valid numbers.

(c) **GFlowNet fine-tuning**
100% of samples
are valid numbers.

Model

R: Reward

q_GFN: 策略网络（由LLM的权重初始化）

T: 序列终止符

Z 1:n: 1到n的序列

$$\mathcal{L}(Z;\theta) = \sum_{0\le i<j\le n} \left(\log \frac{R(z_{1:i}\top) \prod_{k=i+1}^{j} q_{\mathrm{GFN}}(z_k \mid z_{1:k-1}) q_{\mathrm{GFN}}(\top \mid z_{1:j})}{R(z_{1:j}\top) q_{\mathrm{GFN}}(\top \mid z_{1:i})} \right)^2,$$

最终优化目标: $q_{\mathrm{GFN}}^{\top}(Z) \propto R(Z),$

LLM：环境

R：奖励

GFN：策略

Experiment

Sentence Continuation

$$R(Z) = p_{\mathrm{LM}}(Z|X)^{\frac{1}{T}}$$

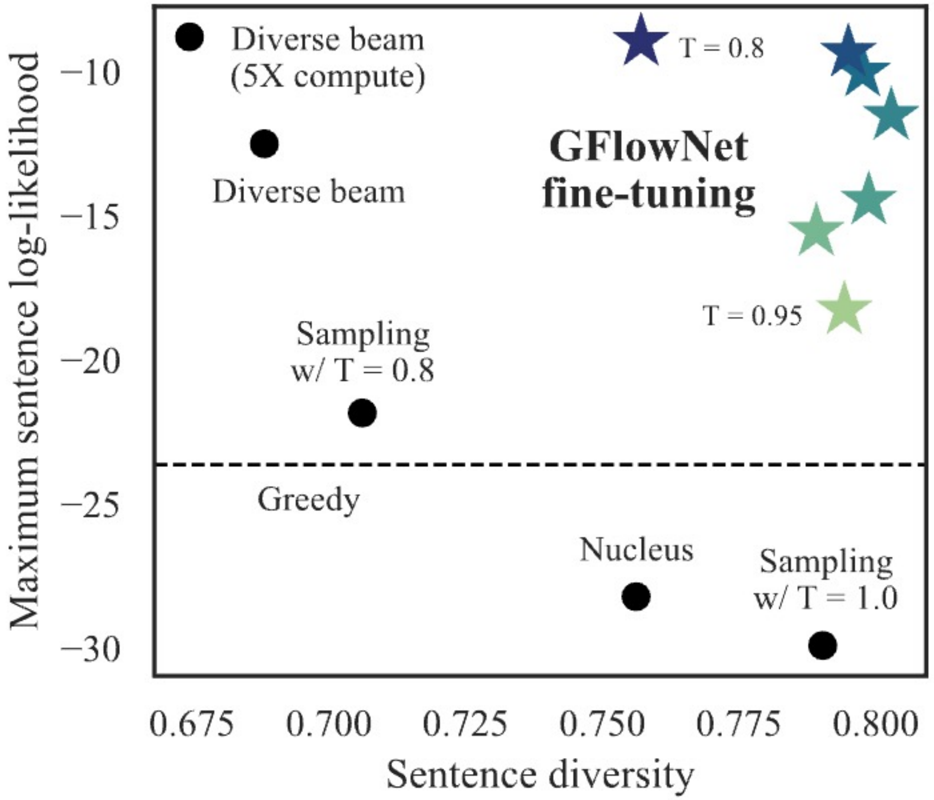


Figure 3: Maximum log-likelihood and diversity of continuations sampled for fixed prompts. GFlowNet fine-tuning (★) samples higher log-likelihood sentences while maintaining more sample diversity than the baselines (● and ---), even when they are given 5× the compute.

Experiment

$$q_{\mathrm{GFN}}(Z|X,Y)$$

Table 4: Test accuracy (%) on an integer arithmetic task with addition and subtraction using a GPT-J 6B model. Training data only include samples with 3 or 4 operands.

Method		Number of Operands		
		In-distribution		OOD
		3	4	5
k-shot CoT	$k = 0$	10.2	6.4	3.2
	$k = 3$	15.8 ± 3.1	11 ± 1.7	5.4 ± 0.2
	$k = 5$	20.4 ± 10.4	17.6 ± 0.6	6.6 ± 1.1
	$k = 10$	26.5 ± 1.4	15.2 ± 1.7	8.9 ± 1.9
	$k = 20$	35.5 ± 1.9	21 ± 1.4	10.5 ± 0.9
Supervised fine-tuning		72.1 ± 1.3	19.6 ± 2.2	12.8 ± 5.7
PPO		30.6 ± 4.1	13.7 ± 4.1	5.6 ± 3.1
GFlowNet fine-tuning		**95.2** ± **1.3**	**75.4** ± **2.9**	**40.7** ± **9.1**

Table 2: Evaluation of the generated infills.

Method	BERTScore	BLEU-4	GLEU-4	GPT4Eval
Prompting	0.081 ± 0.009	1.3 ± 0.5	3.2 ± 0.1	2.4
Supervised fine-tuning	0.094 ± 0.007	1.6 ± 0.8	3.7 ± 0.4	2.7
GFlowNet fine-tuning	**0.184** ± **0.004**	**2.1** ± **0.2**	**4.2** ± **0.7**	**3.4**

Example

Table E.4: Samples generated by PPO fine-tuned and GFlowNet fine-tuned models.

Question (X)	**Generated rationale** (Z)	$\log R$
Question: 1 - 9 + 8 = ? Answer:	1 - 9 - 8	-13.17
	1 - 9 = -8, -8 + 8 = 0	-27.75
Question: 8 + 7 + 2 + 7 = ? Answer:	8 + 7 + 2 + 7	-2.39
	8 + 7 = 15, 15 + 2 = 17, 17 + 7 = 24	-11.72
Question: 7 - 5 + 8 - 0 - 6 =? Answer:	7 - 5 +	-1.22
	7 - 5 = 2, 2 + 8 = 10, 10 - 0 = 10, 10 - 6 = 4	-7.99

Thanks