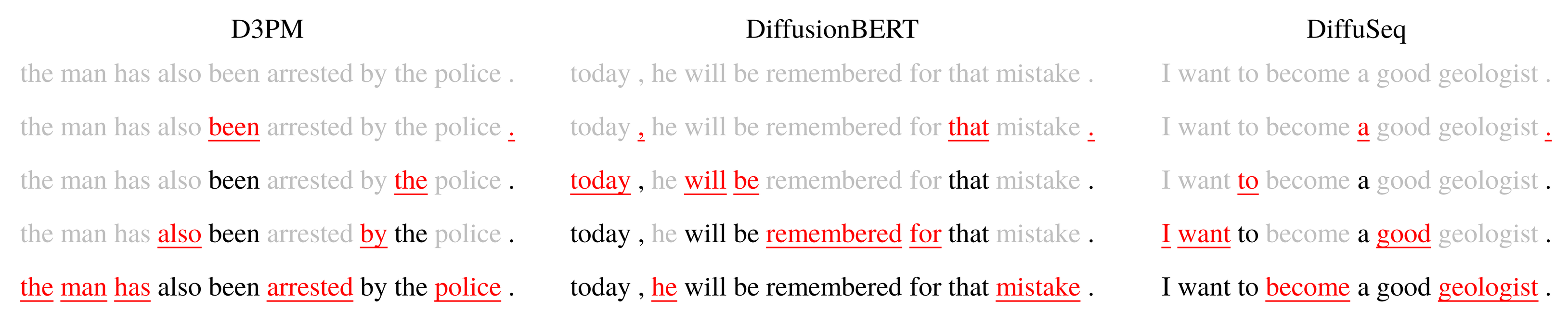
InfoDiffusion: Information Entropy Aware Diffusion Process for Non-Autoregressive Text Generation

Renzhi Wang, Jing Li, Piji Li*

Introduction

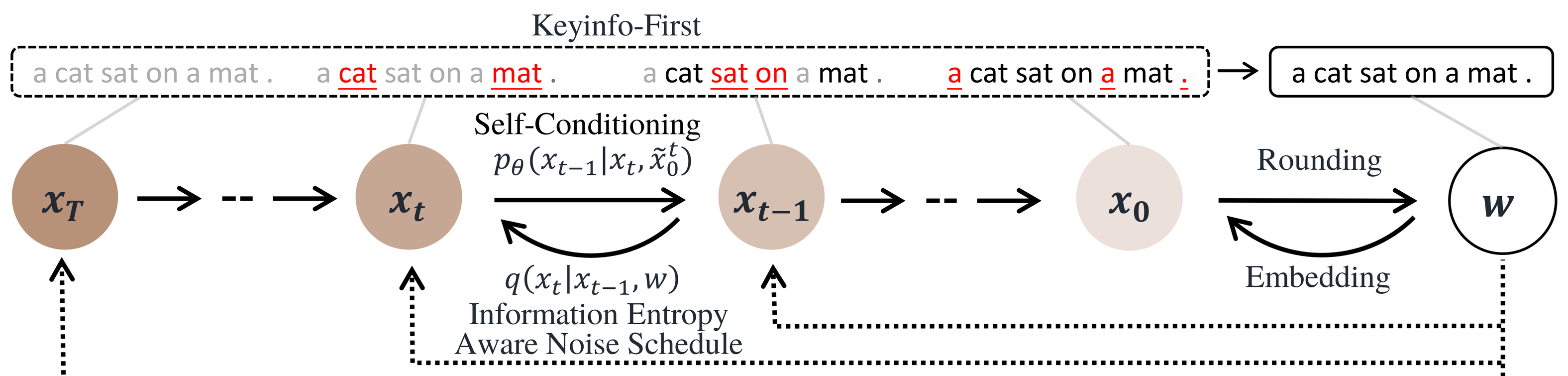


Diffusion models have been increasingly studied for text generation and applied to tasks like named entity recognition and summarization.

There exists a notable disparity between the "easy-first" text generation process of current diffusion models and the "keyword-first" natural text generation process of humans, which could lead to poor generation quality and low efficiency.

To bridge this gap, we propose InfoDiffusion, a non-autoregressive text diffusion model. Our approach introduces a "keyinfo-first" generation strategy and incorporates a noise schedule based on the amount of text information. InfoDiffusion also combines self-conditioning with a partially noising model structure

InfoDiffusion



The overview of the text diffusion model InfoDiffusion. Grey represents undecoded words, red underline indicates words decoded at the current time step, and black represents words decoded in previous time steps.

Noise Schedule

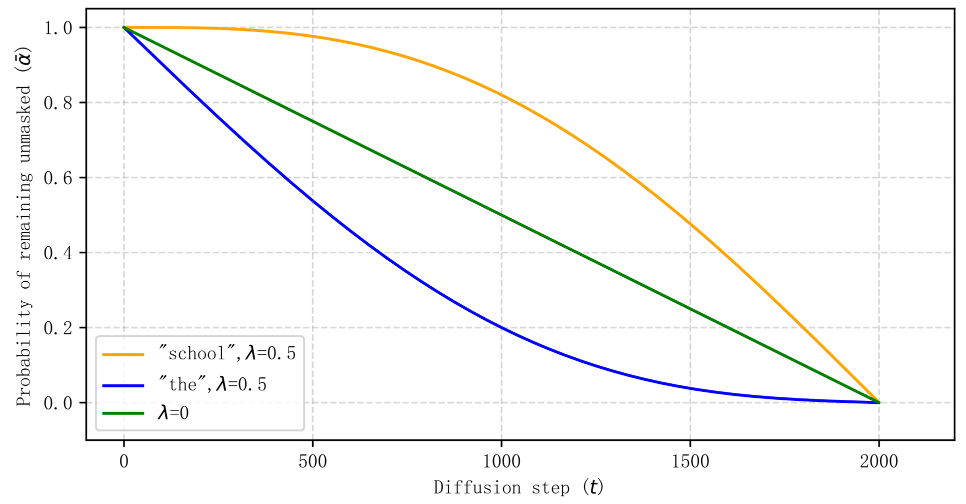
$$q(x_t \mid x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, \sqrt{1-\bar{\alpha}_t}\mathbf{I})$$

$$\bar{\alpha}_t^i = 1 - \frac{t}{T} + \lambda(t)e(w^i) \in [0,1]$$

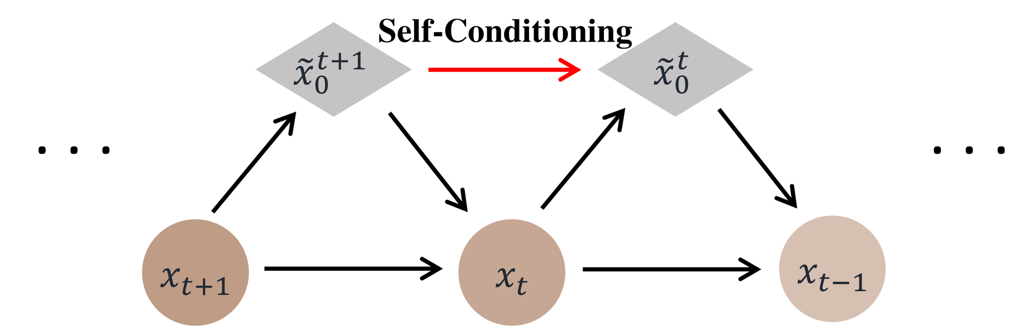
$$\lambda(t) = \lambda \sin(\frac{t}{T}\pi)$$

$$e(w^i) = \frac{H(w^i) - \bar{H}(w)}{max(H(w^j)) - min(H(w^j))}$$

At the initial stage of the forward process, words conveying little meaning are perturbed, while words conveying key information are left intact. Then in the final stage, the key informative words are perturbed. This guides the model to focus first on generating the core semantic content during the reverse process.



Self-Conditioning



An illustration of reverse diffusion sampling steps with Self-Conditioning, sampling directly based on its previously generated samples. The model generates the current prediction $\tilde{x}_0^t(x_t, \tilde{x}_0^{t+1}, t, \theta)$, through a denoising network $f_\theta(x_t, t)$.

Experiments and Results

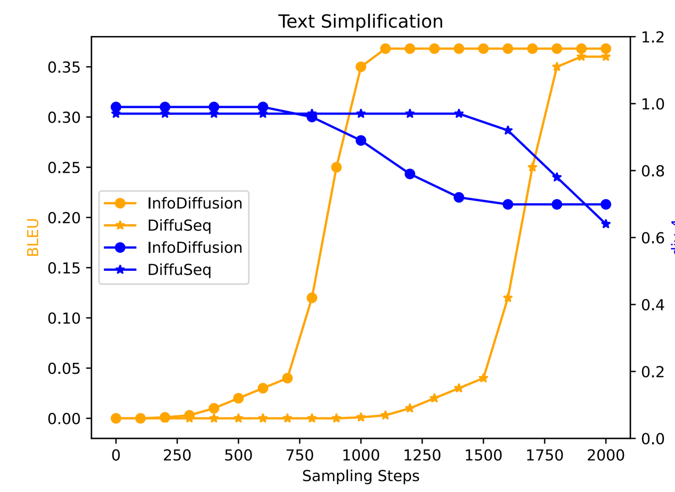
Dataset	Model	Quality			Diversity			Length
		BLEU↑	ROUGE-L↑	BERTScore↑	Dist-1↑	Self-BLEU↓	Diverse-4↑	
Open Domain Dialogue	GRU-attention	0.0068	0.1054	0.4128	0.8998	0.8008	0.1824	4.46
	Transformer-base	**0.0189**	0.1039	0.4781	0.7493	0.3698	0.6472	19.5
	GPT2-base FT	0.0108	**0.1508**	0.5279	0.9194	0.0182	0.9919	16.8
	GPT2-large FT	0.0125	0.1002	**0.5293**	0.9244	0.0213	0.9938	16.8
	GPVAE-T5	0.0110	0.1009	0.4317	0.5625	0.3560	0.5551	20.1
	NAR-LevT	0.0138	0.0550	0.4760	**0.9726**	0.7103	0.1416	4.11
	DiffuSeq	0.0139	0.1056	0.5131	0.9467	**0.0144**	**0.9971**	13.6
	InfoDiffusion	0.0152	0.1272	0.5314	0.9497	0.0152	0.9810	15.3
Question Generation	GRU-attention	0.0651	0.2617	0.5222	0.7930	0.9999	0.3178	10.1
	Transformer-base	0.0364	0.1994	0.5334	0.8236	0.8767	0.4055	12.1
	GPT2-base FT	0.0741	0.2714	0.6052	0.9602	0.1403	**0.9216**	10.0
	GPT2-large FT	0.1110	0.3215	**0.6346**	**0.9670**	0.2910	0.8086	9.96
	GPVAE-T5	0.1251	0.3390	0.6308	0.9381	0.3567	0.7286	11.4
	NAR-LevT	0.0930	0.2893	0.5491	0.8914	0.9830	0.4776	6.93
	DiffuSeq	0.1731	0.3665	0.6123	0.9056	0.2789	0.8103	11.5
	InfoDiffusion	**0.1924**	**0.3892**	0.6310	0.9142	**0.2625**	0.8021	12.7
Text Simplification	GRU-attention	0.3256	0.5602	0.7871	0.8883	0.9998	0.3313	18.9
	Transformer-base	0.2445	0.5058	0.7590	0.8886	0.8632	0.4028	18.5
	GPT2-base FT	0.3085	0.5461	0.8021	0.9439	0.5444	0.6047	16.1
	GPT2-large FT	0.2693	0.5111	0.7882	**0.9464**	0.6042	0.5876	15.4
	GPVAE-T5	0.3392	0.5828	0.8166	0.9308	0.8147	0.4355	18.5
	NAR-LevT	0.2052	0.4402	0.7254	0.9715	0.9907	0.3271	8.31
	DiffuSeq	0.3622	0.5849	0.8126	0.9264	0.4642	0.6604	17.7
	InfoDiffusion	**0.3941**	**0.5997**	**0.8437**	0.9323	**0.4515**	**0.6741**	15.3
Paraphrase	GRU-attention	0.1894	0.5129	0.7763	0.9423	0.9958	0.3287	8.30
	Transformer-base	0.0580	0.2489	0.5392	0.7889	0.7717	0.4312	5.52
	GPT2-base FT	0.1980	0.5212	0.8246	0.9798	0.5480	0.6245	9.67
	GPT2-large FT	0.2059	0.5415	0.8363	**0.9819**	0.7325	0.5020	9.53
	GPVAE-T5	0.2409	0.5886	0.8466	0.9688	0.5604	0.6169	9.60
	NAR-LevT	0.2268	0.5795	0.8344	0.9790	0.9995	0.3329	885
	DiffuSeq	0.2413	0.5880	0.8365	0.9807	**0.2732**	0.8641	11.2
	InfoDiffusion	**0.2656**	**0.5928**	**0.8576**	0.9815	0.2873	**0.8972**	11.4

Analysis

Ablation Study

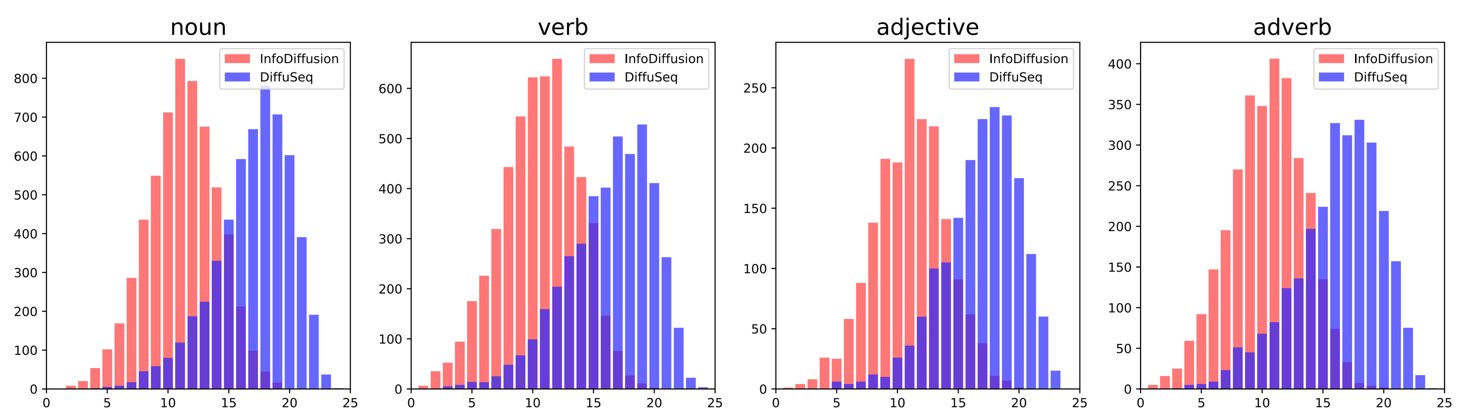
Model	BLEU↑	ROUGE-L↑	BERTScore↑	Dist-1↑
InfoDiffusion	0.2656	0.5928	0.8576	0.9815
- Self-Conditioning	0.2531	0.5884	0.8462	0.9816
- Noise Schedule	0.2480	0.5870	0.8413	0.9798

Inference Efficiency



Case Study

Diffusion Step t	Generation Results of Intermediate Processes $\tilde{x}_0^t$
Input Text	What should i do to be a great geologist?
$t=100$	athan backlash swiped i regentlated patrollingnine jennie ? chill [PAD]
$t=130$	athan backlash swiped i regentlated spotting geologist ? chilean [PAD]
$t=200$	clan patrice swiped i regent carmelgrowth geologist? [unused288] [PAD]
$t=230$	glancing patrice can i heringlated growth geologist ? navigable [PAD]
$t=300$	glance patrice can i moscowgrowth geologist ? corporal [PAD] [PAD]
$t=340$	[CLS] how can i 1859 a 1765 geologist? mcqueen [PAD] [PAD] [PAD]
$t=400$	[CLS] how can i [unused252] a sculpted geologist? [SEP] [PAD] [PAD]
$t=490$	[CLS] how can i 35th a nueva geologist? [SEP] [PAD] [PAD]
$t=600$	[CLS] how can i 35th a sculpted geologist? [SEP] [PAD] [PAD]
$t=840$	[CLS] how can i become a good geologist? [SEP]
$t=950$	[CLS] how can i become a good geologist? [SEP]
$t=1000$	[CLS] how can i become a good geologist? [SEP]
$t=1600$	[CLS] how can i become a good geologist? [SEP]
$t=2000$	[CLS] how can i become a good geologist ? [SEP]



Conclusion

- We propose InfoDiffusion, a novel non-autoregressive text diffusion model, and enables the model to aware the information entropy contained in the text to prioritize generating
- Experimental results demonstrate that InfoDiffusion, which follows a "keyinfo-first" generation order consistent with humans, achieves better generation quality and higher efficiency than baseline models across four text generation tasks.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (No.62106105), the CCF-Baidu Open Fund (No.CCF-Baidu202307), the CCF-Tencent Open Research Fund (No.RAGR20220122), the CCF-Zhipu AI Large Model Fund (No.CCF-Zhipu202315), the Fundamental Research Funds for the Central Universities (No.NJ2023032), the Scientific Research Starting Foundation of Nanjing University of Aeronautics and Astronautics (No.YQR21022), and the High Performance Computing Platform of Nanjing University of Aeronautics and Astronautics.

Code: https://github.com/rzhwang/InfoDiffusion