jh230040

Large-Scale Diffusion Models for Text Generation

Li Zihui (University of Tokyo)

In recent years, diffusion models have exhibited notable advancements in domains such as image and audio processing, emerging as a prominent trend in generative models. However, the discrete signals of textual language in natural language processing (NLP) necessitate further exploration of the application of diffusion models. This study aims to examine the effects and outcomes of the denoising process in diffusion models for text generation tasks involving sequence-to-sequence (Seq2Seq). It investigates the influence of target denoising and full denoising on the performance of paraphrase tasks, while also analyzing the fluctuation patterns of evaluation metrics throughout the training process. Our findings indicate that, in contrast to expectations, full denoising resulted in a decrease in performance for paraphrasing tasks. This decline was observed in both the coherence and fluency of generated text, suggesting that full denoising may be less suited for complex sequence-to-sequence text generation in the current model framework.

1. Basic Information

(1) Collaborating JHPCN Centers

The University of Tokyo
MDX Platform

(2) Theme Area

Data science/data usage area

(3) Research Areas

Very large-scale data processing

(4) Project Members and Their Roles

Main contributors:
Boming Yang, Zihui Li, University of Tokyo;
Aosong Feng, Yale University

Project Consulting:
Yuang Jiang, NEC Lab, USA

Other Members (without significant contributions):
Toyotaro Suzumura, University of Tokyo

2. Purpose and Significance of the Research

Research Purpose:
The goal of this study is to investigate how diffusion models can enhance the text generation capabilities of large-scale, multi-modal pretrained models. Although diffusion models have shown promise in image generation, their application in text generation is not widely studied. This project will focus on integrating diffusion models with existing Transformer structures to improve various text generation tasks, such as text-to-text and image-to-text conversions.

Project Significance:
This research is vital for two reasons. First, it pioneers the use of diffusion models for text generation in NLP, where current models are adept at producing accurate and relevant text, but the incorporation of diffusion models remains untested. Second, it contributes to the ongoing trend of employing large pretrained models in AI, proposing a method to incorporate diffusion models

into these systems efficiently. This approach addresses the significant resource demands of large models, focusing on optimizing efficient inference that benefits both academic research and industrial applications. The project aims to make significant contributions to NLP and AI fields.

3. Significance as JHPCN Joint Research Project

Large pretrained models are garnering significant interest in the field of AI lately. These models require extensive computational power and high-quality data for effective training. Currently, acquiring training data is relatively straightforward, often sourced from open databases or web scraping. However, the real challenge lies in securing sufficient computational resources. The resources available through the JHPCN program could be crucial for advancing this pioneering AI research. Moreover, our research is in line with the Society 5.0 initiative, focusing on the intersection of natural language processing (NLP) and computer vision (CV), two primary AI disciplines. We are hopeful for support from the JHPCN program to drive impactful research in these areas.

4. Outline of Research Achievements up to FY2022 (Only for continuous projects)

5. Details of FY2023 Research Achievements

We have developed an advanced method to demonstrate the impact of the bi-denoising and diffusion processes on the sequence-to-sequence framework. We illustrate the framework in Fig. 1.

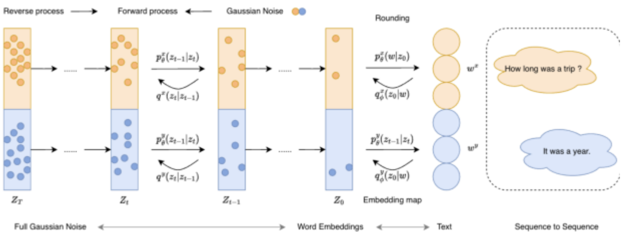


Figure 1: Proposed Framework.

Dual-Noising in the Forward Process: In alignment with DiffusionLM we utilize an embedding layer to convert the discrete text, w, into a continuous space. Notably, in our case, the paired representations (x, y)$are learned concurrently, effectively applying the process to both x and y simultaneously. Following this, a unified embedding is generated from the concatenation of x and y. Upon completion of this process, we are poised to incorporate the standard diffusion forward process into discrete textual input by extending a new Markov transition:

$$q_\phi(z_0|w^{x\oplus y}) = \mathcal{N}(EMB(w^{x\oplus y}), \beta_0 I) .$$

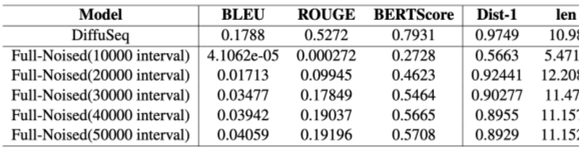
Dual-Denoising in the Reverse Process: The aim of the reverse process is to recover the original z0 by denoising zt:

$$p_\theta(z_{0:T}) := p(z_T) \sum_{t=1}^{T} p_\theta(z_{t-1}|z_t) .$$

To achieve this, we start a series of reverse iterations where each iteration is designed to remove the noise added in the forward process for both x and y. This includes operating an inference network to generate a noise distribution at each time interval. By sequentially applying this network in reverse order, noise is incrementally removed, thereby guiding the combined sequence back to the initial state, z0. The successful recovery of z0 from zt provides further insights into the functional dynamics of the sequence-to-sequence framework, in turn offering potential strategies to intensify the effectiveness of the dual-application denoising procedure. We conducted experiments on the Quota Question Pairs (QQP) dataset, which was collected from the Quora community question answering forum. The dataset consists of 147,000 positive pairs and is used for the Paraphrase Generation Task.

For evaluation, we selected four commonly used metrics: BLEU, ROUGE, and BERTScore. We utilized a Transformer model with 12 layers, a maximum sequence length of 128, an embedding dimension of d = 128, diffusion step T = 2000, and a square-root noise schedule. To mitigate issues with out-of-vocabulary generation, we employed Byte Pair Encoding (BPE) to construct the vocabulary.

The experiments were executed on NVIDIA A100 Tensor Core GPUs, employing four GPUs for training and a single GPU for sampling.



Model	BLEU	ROUGE	BERTScore	Dist-1	len
DiffuSeq	0.1788	0.5272	0.7931	0.9749	10.9
Full-Noised(10000 interval)	4.1062e-05	0.000272	0.2728	0.5663	5.471
Full-Noised(20000 interval)	0.01713	0.09945	0.4623	0.92441	12.20
Full-Noised(30000 interval)	0.03477	0.17849	0.5464	0.90277	11.47
Full-Noised(40000 interval)	0.03942	0.19037	0.5665	0.8955	11.15
Full-Noised(50000 interval)	0.04059	0.19196	0.5708	0.8929	11.15

Table 1: Main result.

Results: As indicated in Tab. 1. we conducted a training process consisting of 50,000 steps and stored checkpoints at every 10,000 steps for decoding purposes. Through a thorough analysis of the variations in evaluation metrics observed at each checkpoint on the test dataset, we investigated the influence of the denoising process employed in the diffusion model.

Initially, at 10,000 steps, all metrics were low as it was still in the early stage of training. By 20,000 steps, the length had stabilized, and the Dist-1 score had risen above 90. In the subsequent 30,000 steps, there was a significant improvement in the BLEU and ROUGE metrics. By 50,000 steps, the model had already learned the most accurate results and answer lengths effectively.

Conclusion: In this study, we explored the use of diffusion processes in both source and target domains for natural language processing tasks. By applying the diffusion forward process to both domains and introducing white noise, we successfully decreased the gap between them, leading them to converge to a similar Gaussian distribution. We then utilized the noised samples from intermediate steps as inputs for a score prediction model. This model's predicted scores enable joint sampling through reverse diffusion processes or conditional sampling by combining forward diffusion in the source with reverse diffusion in the target. Our approach shows significant promise in enhancing performance in language-oriented tasks like QA, machine translation, and dialogue generation. It opens new avenues for the design of coupled diffusion systems for more effective conditional and joint distribution modeling. However, our experimental findings revealed that the full denoising approach, contrary to expectations, led to reduced performance, particularly in complex tasks like machine translation and dialogue generation. This suggests that while the conceptual framework is promising, the current implementation of full denoising requires further refinement to effectively handle the nuances of these language-oriented tasks.

6. Self-review of Current Progress and Future Prospects

Our research on diffusion models in NLP has progressed in line with our initial plan, demonstrating notable findings in text generation. We've successfully investigated the denoising process in diffusion models applied to sequence-to-sequence tasks, focusing particularly on paraphrase generation. Our results intriguingly revealed that full denoising tends to diminish performance, affecting both coherence and fluency of the text. This outcome suggests that current diffusion model frameworks may require adjustments for more complex text generation tasks.

Looking ahead, we plan to expand our experiments on a larger scale. We aim to further explore different denoising techniques and their impact on various NLP tasks beyond paraphrasing. By broadening the scope of our research, we hope to refine our understanding of diffusion models in NLP, specifically in handling the discrete nature of textual language. These future experiments are crucial for advancing the application of diffusion models in text generation and enhancing their effectiveness in more complex, diverse NLP tasks.

7. List of Publications and Presentations

Linyao Chen, Aosong Feng, Boming Yang and Zihui Li. XDLM: Cross-lingual Diffusion Language Model for Machine Translation. Arxiv, 2023 (non-peer reviewed)

Boming Yang, Aosong Feng and Zihui Li. Revisiting Continuous and Discrete Diffusion for Text Generation. Arxiv, 2023 (non-peer reviewed)