Cross-lingual Diffusion Models for Machine Translation

Linyao Chen, Aosong Feng, Irene Li

University of Tokyo, Japan; Yale University, USA

Introduction

🔍 **Significance of the Research:**

- Diffusion models have shown promise in computer vision but require further exploration in natural language processing (NLP).
- Integrating diffusion methods into existing auto-aggressive models in NLP is an open research question.
- Large pretrained models have become a trend and provide a foundation for various tasks, but the high computational cost poses challenges.
- The research aims to integrate diffusion models into pretrained models while optimizing inference speed, benefiting both academia and industry.

💭 **Diffusion Models for Text Generation:**

- Diffusion models have demonstrated strong text generation capabilities in computer vision.
- Research in NLP is needed to explore the application of diffusion models in text generation tasks like summarization and machine translation.
- The challenge lies in integrating diffusion methods into existing auto-aggressive sequence-to-sequence models in NLP.

🚀 **Large Pretrained Models as the New Trend:**

- Recent advancements in large pretrained models, such as GPT models, have shown their effectiveness in various tasks.
- These models are pretrained on massive unlabeled data but require substantial computational resources.
- The research aims to leverage pretrained models while incorporating diffusion models, focusing on efficient integration without starting from scratch to address resource limitations and improve inference speed.

Proposed Methodology

🔬 **Research Theme:** Diffusion Models for Sequence Decoding

- Develop and apply diffusion models to improve sequence decoding in NLP.
- Combine a diffusion model and a decoding model for autoregressive generation.
- Conduct experiments on text summarization and machine translation using large-scale benchmarks.

📊 **Experimental Data:**

- Text summarization: CNN-DM (Daily Mail) news dataset with around 310k articles.
- Machine translation: WMT14 open benchmark with multiple language pairs and millions of sentences.

🔍 **Computational Challenges:**

- Diffusion process and decoding model involve numerous steps, posing computational burden.
- Large-scale datasets and pretrained model initialization further increase computational complexity.
- Estimated requirement of 8 GPUs for this research theme.

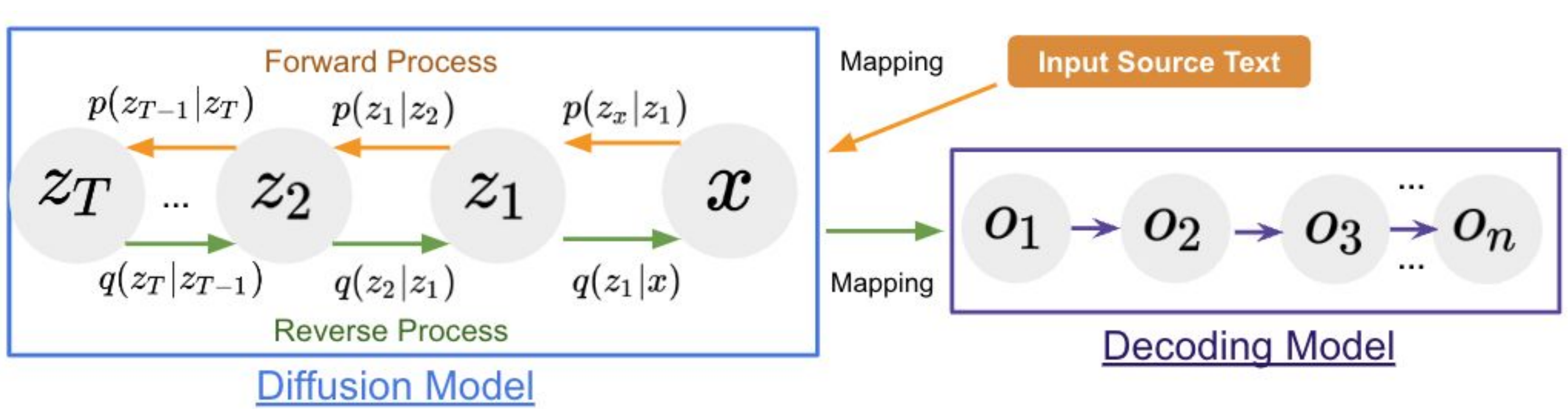

Fig 1.Illustration of the XDiffusion model with sequence decoding.

Expansion to Other Tasks

💡 **Extension to Image-to-Text Generation**

- The encoding process will incorporate image features, and a pretrained model jointly trained on images and texts will be used.
- Evaluation will be conducted on image captioning and visual question-answering tasks using the LAION-5B corpus, with an estimated need for 16 GPUs due to computational resource limitations.

💡 **Computational Challenges and Approach**

- The limited resources necessitate working on a subset of datasets and utilizing existing models rather than training from scratch.
- Completing the training process within weeks is a goal, considering the complexity and scale of the research idea.
- Requesting computational resource support from JHPCN to tackle the computational demands and complexities of the proposed model.

Diffusion Models

Diffusion models: a type of latent variable models trained using variational inference, aim to capture the latent structure of a dataset by modeling the diffusion of data points through the latent space.
In computer vision: utilized to denoise images by reversing the diffusion process, with examples including denoising diffusion probabilistic models, noise conditioned score networks, and stochastic differential equations.
In NLP: training diffusion models that gradually add noise to the input text and then denoise it using a neural network. By learning the diffusion process, these models can generate text or perform other tasks such as text summarization or machine translation.

$$q(\mathbf{x}_{1:T} \mid \mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t \mid \mathbf{x}_{t-1}), \quad q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) := \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}\right)$$

Normal distribution Output Mean Variance

Fig 2.Illustration of the Diffusion Models for Images.

Preliminary Evaluation

Machine Translation with Diffusion Models: we pretrained on Wikidata, and made fair comparison with WMT14 (En to De): Created by Stanford at 2015, the WMT14 English-German Sentence pairs for translation., in Multi-Lingual language. Containing 4.5M in text files.

Strong Baselines:

- Diffuseq (S Gong · 2022)
- Seqdiffuseq (H Yuan · 2022)
- RDM (N Huang · 2023)
- Difformer (Z Gao · 2022)

Preliminary Results

- Evaluated by ROUGE score (eliminate %), our proposed method, XDiffusion is better than original method RDM.

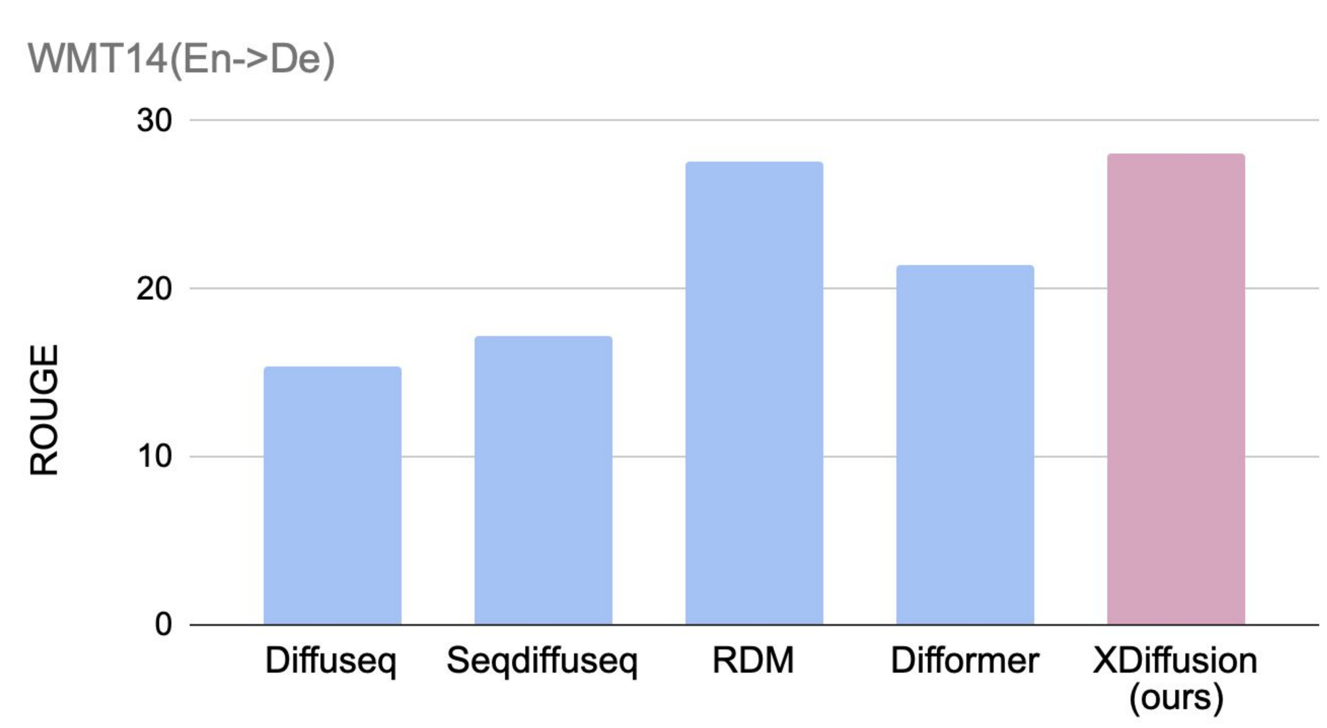

Fig 3.Preliminary Results on WMT14.

Next Steps

Expansion to other languages: we plan to test on other languages, including Japanese and Chinese, and extend our model to be a multilingual one.
Potential for low-resource languages: it is possible to evaluate this model on low-resource languages when the bi-text training data is limited.
Optimization on the efficiency: we hope to optimize on the efficiency by applying approach to speed up the diffusion steps.