SHANSAN GONG

✉ hisansas@gmail.com · ⌂ WebPage: https://summmeer.github.io/

🎓 EDUCATION

Shanghai Jiao Tong University, IE Bachelor (GPA 3.84), CS Master (GPA 3.78) 2015.09 – 2022.03
The University of Hong Kong, CS PhD 2023.09 – Present

👥 RESEARCH EXPERIENCE

My research interests include Diffusion Text Generation, Long Context Language Models.

△ Diffusion Language Models 2022.06 – Present

Previous works that employ diffusion models for text generation tasks have mainly focused on either unconditional text generation or classifier-guided methods, which are not well-suited for Seq2Seq text generation.
First author. Diffuseq: Sequence to Sequence Text Generation with Diffusion Models (ICLR 2023) [code]

- Propose a diffusion model which is designed for Seq2Seq text generation tasks and trained in a classifier-free manner. Also build the connection among AR, NAR, and diffusion models for text generation.

First author. Bridging Discrete and Continuous Text Spaces for Accelerated Diffusion Models (EMNLP 2023)

- We employ state-of-the-art ODE solvers within the continuous space to expedite the sampling process.

Equal first author. Diffusion of Thoughts: CoT Reasoning in Diffusion Language Models (NeurIPS 2024)

- DoT allows reasoning steps to diffuse over time through the diffusion process, offering more flexibility in the trade-off between computation and reasoning performance.

Equal first author. Scaling Diffusion Language Models via Adaptation from Autoregressive Models (preprint)

- We demonstrate connections between AR and diffusion modeling objectives and introduce a simple continual pre-training approach for training diffusion models.

I am interested in further exploring Diffusion Models in text-related generation tasks.

△ Long Context Language Modeling 2022.11 – Present

The ability to process lengthy inputs is crucial for many downstream tasks, including long document QA or dialogue, and scaling up in-context learning examples. Long context language models are still under-explored.
Second author. In-Context Learning with Many Demonstration Examples (pre-print)

- We propose a pre-trained long-range language model and tune it with instructions. We implement incremental encoding and circular position embedding to ensure the extrapolation and efficiency of the model.

Second author. L-Eval: Instituting Standardized Evaluation for Long Context Language Models (ACL 2024)

- We propose a standardized evaluation benchmark. With human-labeled query-response pairs in diverse domains, L-Eval enables a more reliable assessment of long context language models.

△ Information Retrieval - Search and Recommendation 2020.06 – 2022.03

First author. Modeling Implicit Feedback in Session-based News Recommendation (SIGIR 2022)
First author. Transferable and Efficient: Unifying Dynamic Multi-Domain Product Categorization (ACL 2023)

⚙ EXPERIENCE

Shanghai AI Lab Researcher. Diffusion models and long context modeling. 2022.05 – 2023.07
Meituan Research Engineer. Product categorization across domains. 2021.12 – 2022.03
Microsoft STCA NLP Engineer. Query intent recognition of Bing search. 2021.06 – 2021.09

♡ HONORS AND AWARDS

SIGIR 2022 Student Travel Award July. 2022
Outstanding Graduate in Shanghai Municipality Mar. 2022