Jongyoon Song

coms1580@gmail.com • +82 (10) 2617-2038 • LinkedIn

RESEARCH INTEREST

My research interest lies in implementing reliable generative (large) language models. I focus on analyzing issues such as knowledge conflict and hallucination in language models and improving their human alignment. Furthermore, I am interested in the overall research topic related to virtual assistants capable of human-like conversations across various domains.

- **Keywords**: Generative (large) language model, knowledge conflict & hallucination, trustworthy AI

INTERNSHIP

Microsoft Research Asia (MSRA) Dec 2018 – Mar 2019

- Improving Text-to-speech Model by Mel-spectrogram Pre-training
  - Pre-train the bi-directional mel-spectrogram model using the modified objectives of BERT, and utilize the model a module within an encoder-decoder based text-to-speech model.

Kakao Enterprise Feb 2021 – Jun 2021

- Non-autoregressive Neural Machine Translation
  - Propose alignment decomposition method to improve word alignment estimation accuracy and coherence of non-autoregressive translation.

EDUCATION

Seoul National University (SNU) Mar 2011 – Feb 2016

- B.S. in Electrical and Computer Engineering
- Cumulative GPA: 3.53/4.3

Seoul National University (SNU) Mar 2017 – present

- In PhD course in Electrical and Computer Engineering
- Cumulative GPA: 3.83/4.3

HONORS & AWARDS

Best Paper Award, ISSIP-IBM Smart Service System Best Paper Award (2018)

SCHOLARSHIP

National Science & Technology Scholarship, Korean Student Aid Foundation, 2011 – 2016

PUBLICATIONS

[1] J. Song, J. Lee, H. Kim, E. Choi, M. Kim, S. Yoon "Customization of IBM Intu's Voice by Connecting Text-to-Speech Services with a Voice Conversion Network," *Hawaii International Conference on System Sciences (**HICSS**)*, Jan 2018.

[2] S. Kim, S. Lee, J. Song, J. Kim, S. Yoon "FloWaveNet : A Generative Flow for Raw Audio," *International Conference on Machine Learning (**ICML**)*, Jun 2019.

[3] J. Song, S. Kim, S. Yoon "AligNART: Non-autoregressive Neural Machine Translation by Jointly Learning to Estimate Alignment and Translate," *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (**EMNLP**)*, **Oral**, Nov 2021.

[4] S. Yu, J. Song, H. Kim, S. Lee, W. Ryu, S. Yoon "Rare Tokens Degenerate All Tokens: Improving Neural Text Generation via Adaptive Gradient Gating for Rare Token Embeddings," *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (**ACL**)*, Mar 2022.

[5] J. Song, N. Park, B. Hwang, J. Yun, S. Joe, Y. Gwon, S. Yoon "Model Intrinsic Features of Fine-tuning based Text Summarization Models for Factual Consistency," *Findings of the 61th Annual Meeting of the Association for Computational Linguistics (**ACL**)*, Jul 2023.

[6] J. Song, N. Park, B. Hwang, J. Yun, S. Joe, Y. Gwon, S. Yoon "Entity-level Factual Adaptiveness of Fine-tuning based Abstractive Summarization Models," *the 18th Conference of the European Chapter of the Association for Computational Linguistics (**EACL**)*, Mar 2024.

RESEARCH EXPERIENCE

Graduation Thesis Jul 2016 – Feb 2017

- Data Organization Methods for Efficient Learning of Watson's Question Answering Model
  - Analyze relation between data organization for question answering model and performance changing scenarios, for example, varying the ratio of yes/no and wh- question or paraphrasing questions.

Non-autoregressive Neural Machine Translation May 2020 – Nov 2020

- Non-autoregressive Neural Machine Translation by Learning to Estimate Alignment and Translate
  - Leverage alignment information to alleviate multi-modality problem of non-autoregressive neural machine translation models, while the encoder effectively learns to estimate alignment.

Abstractive Text Summarization Mar 2021 – Feb 2024

- Abstractive Text Summarization with Enhanced Factual Consistency
  - Probe the cause of low factual consistency of abstractive text summarization model which is fine-tuned from pre-trained language model and improve the factual consistency during fine-tuning.

Large Language Models Feb 2024 – present

- Hallucination Problem in Large Language Models
  - Identify the various types of hallucinations and analyze the inductive biases in supervised fine-tuning and human alignment that cause the hallucinations.

- Analysis and Optimization of Instruction and Preference Data
  - Research on data-efficient human alignment methods by analyzing instruction and preference data.

PROJECTS

Patent Classification (Hyundai Motor Company) Jan 2020 – Dec 2020

- Improving Classification Accuracy on the Long-tail Patent Dataset
  - Utilize re-labeling, pre-trained weight re-initialization, and two-stage fine-tuning to improve patent classification accuracy.
  - Leverage attention weights to extract keywords in the patent classification process.

Addressing Shortcut Problem of Masked Language Models (LG AI Research) Jun 2021 – May 2022

- Global Context Learning Using Improved Masked Language Modeling
  - Research on optimal masking strategy during the masked language modeling process to prevent the model from learning shortcut that interferes global context learning.

Factually Consistent Text Summarization Models (Samsung SDS) Aug 2021 – Jun 2023

- Model Intrinsic Feature Analysis of Fine-tuning based Text Summarization Models
  - Analyze various text summarization models that are proposed to improve factual consistency and find that the distribution of summary in the fine-tuning set affects factual consistency.

- Improving Factual Adaptiveness of Fine-tuning based Text Summarization Models
  - Analyze the robustness of summarization models to knowledge conflict and propose a counterfactual data augmentation approach that incorporates contrastive learning techniques to improve factual consistency in cases of knowledge conflict.

INVITED TALKS

"(Relative) Inductive Bias for Various Problem", Naver, 2018
"Factual Consistency and Adaptiveness for Abstractive Summarization Models", Samsung SDS, 2023

REFERENCES

- **Sungroh Yoon**
  Professor
  Electrical and Computer Engineering
  Seoul National University
  sryoon@snu.ac.kr • +82 (2) 880-1406