Ahjeong **Park**

AI RESEARCHER · NLP ENGINEER

Seoul, Rep. of KOREA

ajeongi59@gmail.com | aaajeong | aaajeong | ajeongi59 | azeomi

"Done is better than perfect"

Summary

I am a graduate M.S. student in the School of IT Engineering at Sookmyung Women's University co-advised by Prof. Chulyun Kim. I received my M.S. in IT Engineering from Sookmyung, where I worked with Prof. Chulyun Kim.

My primary research interests lie in the area of natural language processing (NLP) and Ensemble in Generative Models. The long-term goal of my research is to enhance the practicality of NLP systems (e.g., chatbot, nmt, AI assistant) so that they can be widely used in real-world scenarios. In particular, my research focuses on 1) How NLP models collaborate with each other efficiently (i.e. ensemble, and 2) the ability of each model to cooperate with each other by answering what they can do well and by answering that they do not know what areas of high uncertainty are high. In addition, 3) technology for automatically generating natural language annotations from programming code.

Education

Sookmyung Women's University — *Seoul, S.Korea*

B.S. in IT Engineering — *Mar. 2016 - 2. 2021*

Sookmyung Women's University — *Seoul, S.Korea*

M.S. in IT Engineering — *Mar. 2021 - 2. 2023*

- Got the outstanding Alumni Scholarship, which is given to promising students from our undergraduate school.

Honors & Awards

PATENTS

2023	**Lead**, Method and system for ensemble of recurrent neural network model	*International*
2023	**Lead**, Method and apparatus for automatically generating natural language comments based on transformer	*International*
2022	**Lead**, Method and system for ensemble of recurrent neural network model	*Domestic*
2022	**Lead**, Method and apparatus for automatically generating natural language comments based on transformer	*Domestic*

AWARDS

2019	**Grand Prize**, 2019 Public SW Contribuithon	*Seoul, S.Korea*
2018	**1st Place**, The 4th Global Innovator Festival (Makerthon)	*Seoul, S.Korea*
2018	**3rd Place**, AWS Women in Tech Hacking Competition Final	*Seoul, S.Korea*

Publications

REGEN: Recurrent Ensemble Methods for Generative Models — *Masters' Thesis*

Ahjeong Park & Chulyun Kim — *2022*

- Contributed as the first author in my master's thesis, taking on the responsibilities of conducting experiments and overall paper writing.

ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information — *IEEE Access*

Youngmi Park & Ahjeong Park & Chulyun Kim — *2023*

- Served as a hands-on practitioner for a government project, contributing as the second author to research.
  My responsibilities included designing the CAT dataset and contributing to modeling, as well as participating in the paper-writing process

A Study on the AST Traversal Method to Improve the Quality of Code Comment Generation — *KCC*

Youngmi Park & Ahjeong Park & Chulyun Kim — *2022*

- Participated as the second author in an academic conference presentation paper, where I was responsible for conducting experiments and paper writing.

Improving Seq2Seq model performance through new RNN ensemble method

KSC

Ahjeong Park & Chulyun Kim

2021

- Took a lead role as the first author in an academic conference presentation paper, being responsible for conducting experiments and paper writing.

Research

A new ensemble algorithm for natural language generation and translation models(REGEN)

Seoul, S.Korea

NLP & Generation & Ensemble

Jul. 2021 - Dec. 2022

1. **Problem**
   - The conventional ensemble treats deep learning models as black boxes, aggregating only the final results, which poses limitations in terms of length, scope, and suitability for high-diversity generative models.
2. **Contribution**
   - We propose a new ensemble called REGEN, which is suitable for generative tasks, involving the iterative prediction of the next word.
   - REGEN is composed of Seq2Seq and Transformer models, where each model's decoder influences the next generation at each step, and we introduce algorithms for model agreement called Consensus and Survival Ensemble.
3. **Methods**
   - Seq2Seq, Transformer, Bagging, Voting(Soft voting, Hard voting)
4. **Results**
   - Performance Evaluation: We conducted machine translation experiments as the generative task, and it was confirmed that REGEN outperforms not only single-component models but also traditional ensembles.
   - Application of Seq2Seq model: Single (19.56 BLEU) < Traditional Ensemble (20.64 BLEU) < REGEN (22.71 BLEU)
   - Application of Transformer model: Single (24.93 BLEU) < Traditional Ensemble (23.76 BLEU) < REGEN (25.38 BLEU)
5. **Applicable Model**
   - Auto-Regressive Model(GPT etc), Encoder-Decoder Model
6. **Achievement**
   - Master's Thesis
   - Domestic and International Patent Applications
   - Preparing for IEEE Access Journal Submission

Automatically Generating Natural Language Comments for Deep Learning-Based Source Code

Seoul, S.Korea

NLP & Auto Code Comment Generation

Jan. 2022 - Jun. 2023

1. **Introduction**
   - Participating as a collaborative research and development institution in the Korea Creative Content Agency Government Research Project.
   - Project Title: "Development of Copyright Application Technology for Supporting Fair Software Copyright Transactions and Distribution Ecosystem.
   - The need for automatic natural language comment generation for source code has increased as the scale of software grows.
2. **Problem**
   - There are limitations in the existing source code comment generation due to the loss of syntactic information and the lack of alignment with the source code.
3. **Contribution**
   - We propose a novel approach combining the ALSI-Transformer, a deep learning-based model for automatically generating natural language comments for source code, and CAT (Code-Aligned Type), which provides syntactic information for the source code.
4. **Methods**
   - Transformer, CNN, AST(Abstract Syntax Tree), Model Compression
5. **Results**
   - By addressing the limitations of existing datasets for source code comment generation and improving the model architecture, we have achieved a state-of-the-art (SOTA) model for accurate and compact comment generation tasks.
6. **What did I do**
   - Conducted research investigation on comment generation and analyzed the state-of-the-art (SOTA) models.
   - Designed and implemented the new data type CAT (Code-Aligned Type) and successfully extracted a total of 465,812 data.
   - Utilized lightweight model algorithms (CNN filter & Gate Network) to reduce the model size by approximately 12%.
   - Designed and experimented with six methods to combine multi-modal (Lexical & Syntactic) aspects of source code into a single encoder, leading to the final ALSI-Transformer model with improved performance
     - Average (50.75 BLEU) →Concat (51.41 BLEU) →Addition (53.21 BLEU) →Gate Network (53.80 BLEU).
   - Compressed the model size from 10 GB to 8.3 GB by converting the previous SOTA model (2-encoder) into a 1-encoder model.
   - Conducted comparative experiments with baselines, confirming the SOTA performance
     - Baseline (49 BLEU) →ALSI-Transformer (53.80 BLEU).
   - Paper writing
7. **Achievement**
   - IEEE Access, 2023
   - Domestic and International Patent Applications

Extracurricular Activity

7th PseudoLab - Summarizing the Latest Research Trends in an Engaging and Accessible Manner

Seoul, S.Korea

Core Member & Learner — *Sep. 2023 - Nov. 2023*

- A study group dedicated to summarizing the latest research trends since 2022 from our unique perspectives.
- Our mission is to provide engaging summaries of current research developments that even non-specialists can enjoy and understand.
- Sharing the Latest Research Trends in the NLP Field, with a Focus on LLM Fine-Tuning and Parameter-Efficient Fine-Tuning.

NLP Paper Study Group

Seoul, S.Korea

Core Member — *Aug. 2022 - Nov. 2022*

- A Study Group for Reading Important Papers in Natural Language Processing, Conducting Paper Reviews, Implementing Code, and Recording Presentations.
- In Charge of Overseeing the Study Progress and Summarizing Key Papers in Natural Language Processing.
- Sharing Paper Study Presentation Materials
- Uploading Natural Language Processing Paper Study Videos on YouTube