

박아정 (Ahjeong Park)

Graduated M.S. Student at Sookmyung Women's Univ

ahjeong@sookmyung.ac.kr | [Linked In](#) | [Github](#) | [Blog](#)

RESEARCH INTERESTS

Natural Language Processing & Ensemble

The long-term goal of my research is to enhance interaction with AI using NLP systems so that they can be widely used in real-world. In particular, my research focuses on ensemble methods that can develop NLP models that make humans comfortable.

EDUCATION

Sookmyung Women's University

M.S. in IT Engineering Mar. 2021 - Feb. 2023

- Research Assistant at Knowledge and Information Engineering Lab (Advisor: Dr. Chulyun Kim)
- GPA: 4.30 / 4.30 (4.50 / 4.50)

Sookmyung Women's University

B.E. in IT Engineering Mar. 2016 - Feb. 2021

- GPA: 3.51 / 4.30 (3.81 / 4.50)

PUBLICATIONS

REGEN: Recurrent Ensemble Methods for Generative Models (투고 중)

2023, IEEE Access 준비 중

Ahjeong Park, Youngmi Park, Chulyun Kim

ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information (심사 중)

2023, IEEE Access 준비 중

Youngmi Park, **Ahjeong Park**, Chulyun Kim

코드 주석 생성 품질 개선을 위한 AST 순회 방법에 관한 연구

2022 한국컴퓨터종합학술대회

박영미, **박아정**, 김철연

RNN의 새로운 앙상블 기법을 통한 Seq2Seq 모델 성능 개선

2021 한국소프트웨어종합학술대회

박아정, 김철연

PATENT

재귀 신경망 모델의 앙상블 방법 및 시스템

출원일자: 2022년 12월 12일

출원인: 숙명여자대학교 산학협력단

발명자: 김철연, **박아정**, 박영미

트랜스포머 기반의 자연어 주석 자동 생성 방법 및 장치

출원일자: 2022년 12월 12일

출원인: 숙명여자대학교 산학협력단

발명자: 김철연, 박영미, **박아정**

EXPERIENCES

- 주)라이온브리지테크놀로지스코리아
근무부서: Lionbridge AI
프로젝트: Samsung partnership Bixby Project
 - Bixby 연락처(Contact) Capsule 개발 및 유지보수
 - Bixby 학습 데이터 관리 및 생성
 - OS & Device 에 따른 Bixby UX/UI 가이드 변경 및 모델 수정
 - 아태여성정보통신원, UNESCO-UNITWIN 인도네시아 현지교육 ICT&LearderShip, ICT 교육
활동장소: 인도네시아 Universitas Gadjamada(Yogyakarta 소재)
[강의보조]
 - 전반적인 강의 진행 및 보조 및 강사 요청 사항 협조
 - ICT 실습 조교 리더(아두이노, 모바일 앱 개발)
[현지 학생 관리]
 - 팀의 형태로 1인당 8명의 학생 배정
 - 교육 기간 내 출석 체크 및 적극적인 참여 독려
 - SMWU-Kyushu Univ 1st Interdisciplinary Hackathon Program
활동장소: 일본 규슈
 - 한국인 4명, 일본인 1명과 팀을 이루어 일본의 사회적 문제를 해결하는 IoT 프로젝트를 진행

Nov. 2019 ~ Jul. 2020

Aug 11. 2019 ~ Aug 17. 2019

Feb 10. 2019 ~ Feb 19. 2019

AWARDS

- 2019 공개SW 컨트리뷰톤 ‘최우수상(정보통신산업진흥원장상)
 - 제 4회 글로벌 이노베이터 페스타(메이커톤) 1등 ‘교육부장관상
 - AWS Women in Tech Hackathon 3등 ‘한국여성과학기술단체총연합회장상

2019

2018

2018

PROJECTS

- 공정한 SW저작권 거래 및 유통 생태계 지원을 위한 저작권 응용 기술 개발 - 자연어 주석 생성 연구 ([상세설명](#)) 2022
- 2021 NH 투자증권 빅데이터 경진대회 - 주식 보유기간 예측 및 서비스 아이디어 제안 ([상세설명](#)) 2021
- 2019 공개 SW 컨트리뷰톤 - You Only Look Keras 오픈소스 프로젝트 ([상세설명](#)) 2019
- 학부 졸업 프로젝트 - 추카(Chuka), 이미지 인식을 이용한 축구 하이라이트 영상 추출 프로그램 ([상세설명](#)) 2019
- Smart House(음성인식 스마트 하우스) ([상세설명](#)) 2019
- 프리맘(Pre-Mom) - 임신부의 편리한 지하철 이용을 위한 서비스로 핑크카펫 좌석 알람 및 핑크라이트 IoT 서비스 ([상세설명](#)) 2018

1. 석사논문: Recurrent Ensemble Methods to Enhance the Performance of Multiple Recurrent Network Models

요약

기존의 앙상블은 앙상블 단계에서 오직 모델의 최종 output 결과만을 집계합니다. 이는 최종 output의 diversity가 낮은 모델에서는 적절합니다. 하지만 Generative 계열의 모델은 최종 output의 길이와 범위에 제한이 없고 diversity가 높습니다. 이러한 점은 각 모델의 최종 output 간 합의를 보는데 문제가 있습니다. 따라서 이에 적합한 새로운 Recurrent Ensemble을 제안하고 다양한 실험을 통해 기존의 앙상블 방법보다 성능이 향상됨을 확인합니다.

실험

대표적으로 Seq2Seq, Transformer 모델을 기반으로 기계번역 등의 실험을 진행했습니다. 2가지의 Recurrent Ensemble로 Consensus & Survival Ensemble을 제안했고 이와 비교하기 위해 기존의 앙상블을 새롭게 해석한 Majority Ensemble을 설계했습니다. 각 앙상블은 2개에 모델에서 동일한 과정으로 동작합니다.

모델 별 앙상블 그림: Seq2Seq

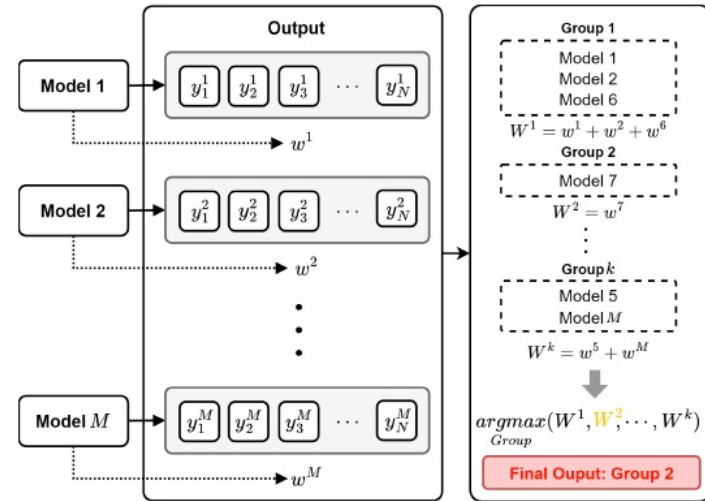


Figure 3.1: Overall Structure of Traditional Majority in Seq2Seq

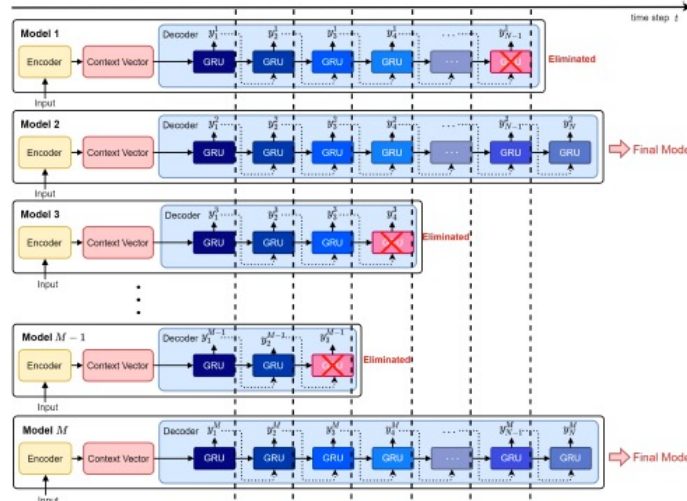


Figure 3.3: Overall Structure of Survival in Seq2Seq

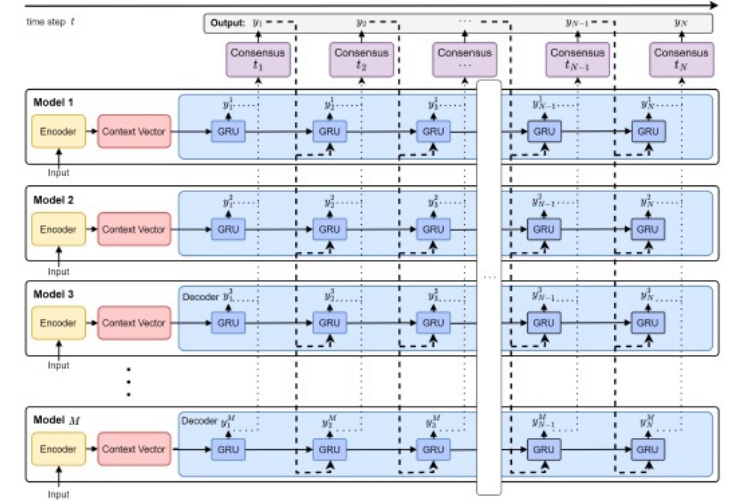


Figure 3.5: Overall Structure of Consensus in Seq2Seq

1. 석사논문: Recurrent Ensemble Methods to Enhance the Performance of Multiple Recurrent Network Models

모델 별 앙상블 그림: Transformer

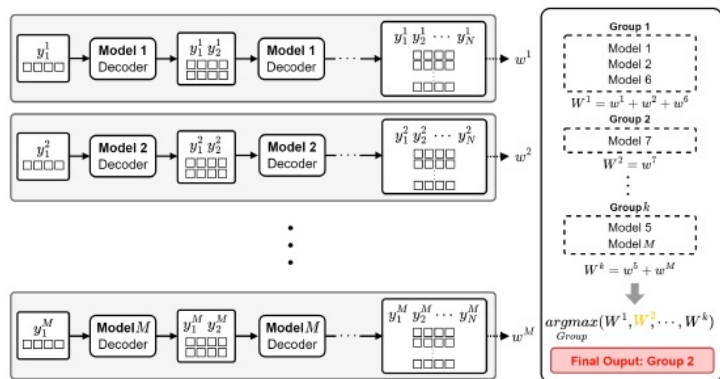


Figure 3.7: Overall Structure of Traditional Majority in Transformer

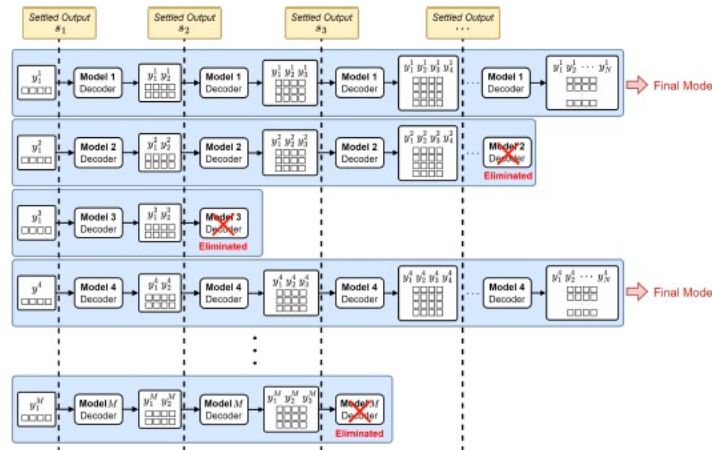


Figure 3.9: Overall Structure of Survival in Transformer

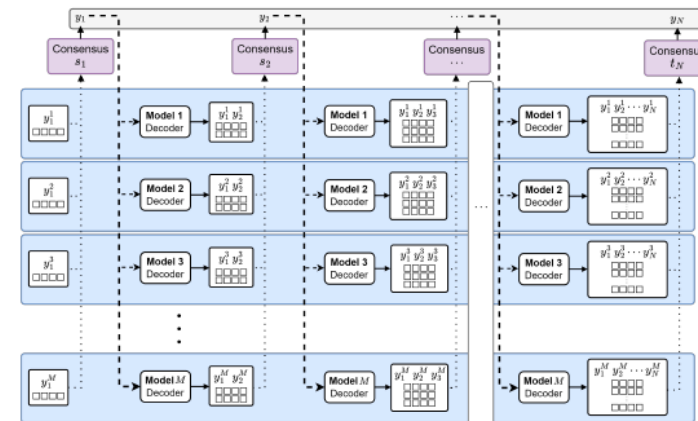


Figure 3.11: Overall Structure of Consensus in Transformer

실험결과

Seq2Seq, Transformer 모델 단일 모델과 기존 앙상블 방법과 비교했을 때 좋은 성능을 확인했습니다. 따라서 새롭게 제안한 Recurrent Ensemble이 효과적임을 입증했습니다.

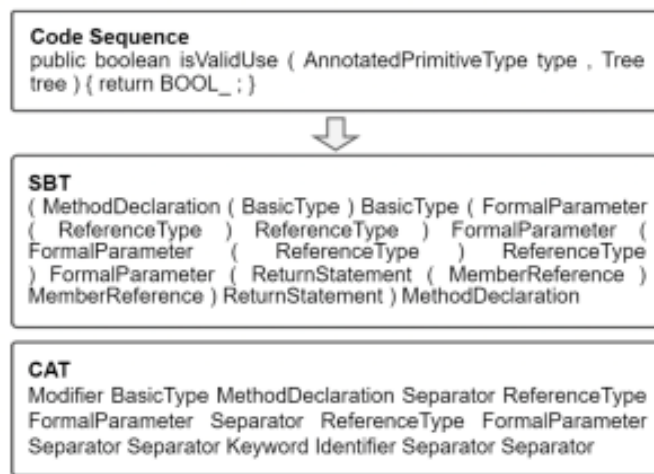
2. ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information (심사중)

요약

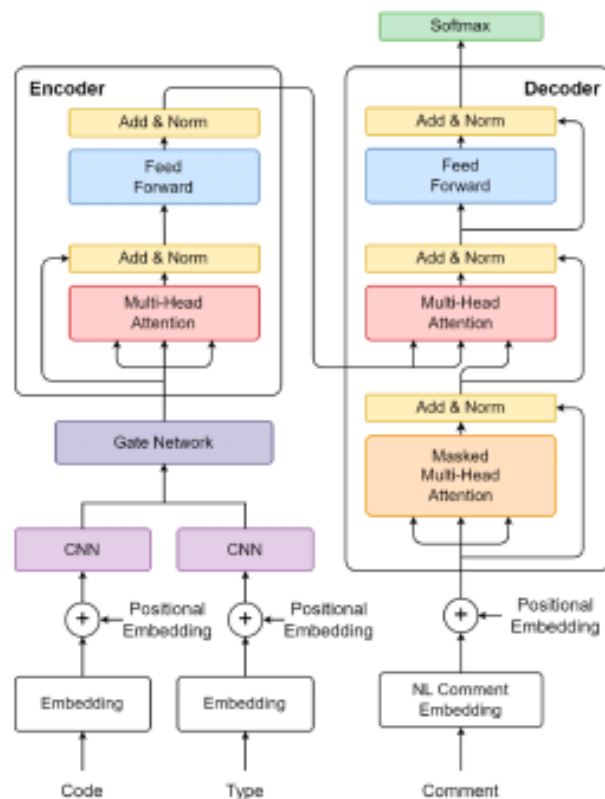
코드 주석은 컴퓨터 프로그램의 운영 과정을 설명하고 디버깅과 유지보수와 같은 프로그래밍 작업의 장기적인 생산성을 증가시킵니다.

따라서 프로그래밍 코드에서 자연어 주석을 자동으로 생성하는 기술은 필수적입니다.

본 논문에서는 자동 주석 생성의 정확도를 향상시키기 위해 어휘 및 구문 정보의 순서와 길이를 정렬하고 새로운 구문 시퀀스인 코드 정렬 유형 시퀀스(CAT)를 제안했습니다. 트랜스포머를 기반으로 새로운 신경망 모델인 ALSI-Transformer를 제안합니다. 다양한 실험을 통해 코드 주석 생성에서 좋은 성능을 달성함을 입증했습니다.



Code-Aligned Type sequence(CAT)



ALSI-Transformer 모델 구조