

도전과 융합을 좋아하는

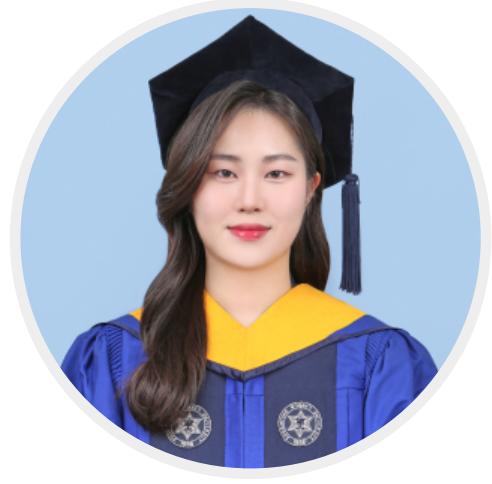
연구자 박아정입니다

Potfolio

CONTACT

ahjeong@sookmyung.ac.kr
010 7448 8798





도전과 융합을 좋아하는
연구자
박아정입니다

RESEARCH INTERESTS

Natural Language Processing(NLP) & Ensemble in Deep Learning

My primary research interests lie in the area of natural language processing (NLP) and Ensemble in Deep learning. The long-term goal of my research is to enhance the practicality of NLP systems (e.g., neural machine translation) so that they can be widely used in real-world scenarios.

Keywords

- Ensemble in Deep Learning
- Neural Machine Translation
- Automatic Code Comment Generation
- Generative Model in NLP
- Efficiency

박아정 / Ahjeong Park

1997.05.09 / 서울특별시

Tel. 010-7448-8798

Email. ahjeong@sookmyung.ac.kr

서울특별시 은평구 역촌동

[Linked In](#)

[Github](#)

[Blog](#)

GRADUATION

Sookmyung Women's University

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

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

Sookmyung Women's University

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

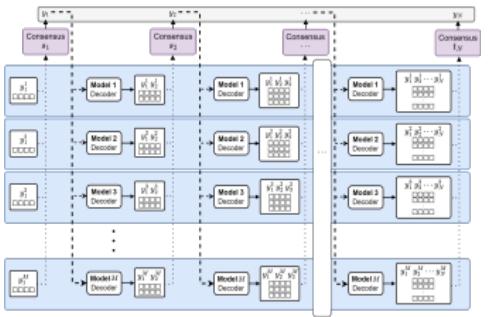
- GPA: 3.51 / 4.30 (3.81 / 4.50)

Daejeon Girl's High School

Mar. 2013 - Feb. 2016

- Student president

PUBLICATIONS

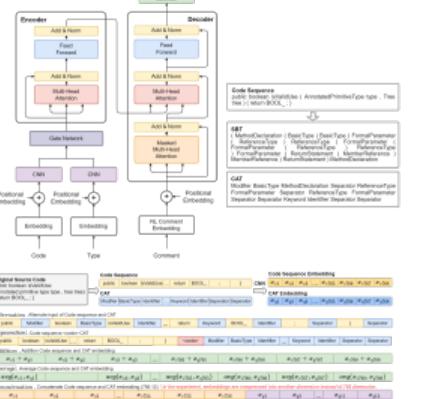


◀ REGEN: Recurrent Ensemble Methods for Generative Models (투고 준비 중)

Ahjeong Park, Youngmi Park, Chulyun Kim

2023, IEEE Access 준비 중

Slides

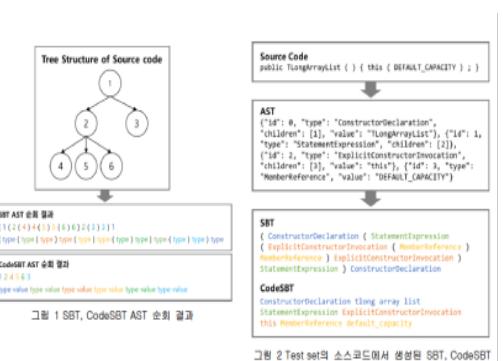


◀ ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information (심사 중)

Youngmi Park, Ahjeong Park, Chulyun Kim

2023, IEEE Access

Slides



◀ 코드 주석 생성 품질 개선을 위한 AST 순회 방법에 관한 연구

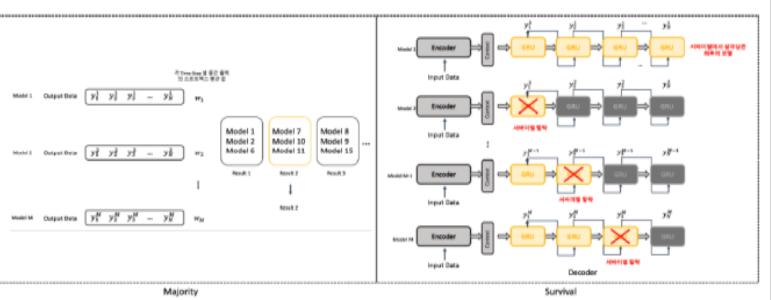
박영미, 박아정, 김철연

2022 한국컴퓨터종합학술대회

Paper

Poster

Code



◀ RNN의 새로운 양상을 기법을 통한 Seq2Seq 모델 성능 개선

박아정, 김철연

2021 한국소프트웨어종합학술대회

Paper

Poster

Code

PATENT

재귀 신경망 모델의 양상을 방법 및 시스템

심사중

출원일자: 2022년 12월 12일

출원인: 숙명여자대학교 산학협력단발명자: 김철연, **박아정**, 박영미

트랜스포머 기반의 자연어 주석 자동 생성 방법 및 장치

심사중

출원일자: 2022년 12월 12일

출원인: 숙명여자대학교 산학협력단발명자: 김철연, 박영미, **박아정**

AWARDS

- 2019 공개SW 컨트리뷰톤 ‘최우수상(정보통신산업진흥원장상)’ 2019
- 제 4회 글로벌 이노베이터 페스타(메이커톤) 1등 ‘교육부장관상’ 2018
- AWS Women in Tech Hackathon 3등 ‘한국여성과학기술단체총연합회장상’ 2018

EXPERIENCES

- **주)라이온브리지테크놀로지스코리아**

Nov. 2019 ~ Jul. 2020

근무부서: Lionbridge AI

프로젝트: Samsung partnership Bixby Project

- Bixby 연락처(Contact) Capsule 개발 및 유지보수
- Bixby 학습 데이터 관리 및 생성
- OS & Device에 따른 Bixby UX/UI 가이드 변경 및 모델 수정

- **아태여성정보통신원, UNESCO-UNITWIN 인도네시아 현지교육 ICT&LearderShip, ICT 교육**

Aug 11. 2019 ~ Aug 17. 2019

활동장소: 인도네시아 Universitas GadjahMada(Yogyakarata 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**

Feb 10. 2019 ~ Feb 19. 2019

활동장소: 일본 규슈

- 한국인 4명, 일본인 1명과 팀을 이루어 일본의 사회적 문제를 해결하는 IoT 프로젝트를 진행

PROJECTS

- 공정한 SW 저작권 거래 및 유통 생태계 지원을 위한 저작권 응용 기술 개발 - 자연어 주석 생성 연구 Slides 0 2022
- 2021 NH 투자증권 빅데이터 경진대회 - 주식 보유기간 예측 및 서비스 아이디어 제안 Slides 0 2022
- 2019 공개 SW 컨트리뷰튼 - You Only Look Keras 오픈소스 프로젝트 Slides 0 2022
- 학부 졸업 프로젝트 - 추카(Chuka), 이미지 인식을 이용한 축구 하이라이트 영상 추출 프로그램 Slides 0 2022
- Smart House(음성인식 스마트 하우스) Slides 0 2022
- 프리맘(Pre-Mom) - 임신부의 편리한 지하철 이용을 위한 서비스로 핑크카펫 좌석 알림 및 핑크라이트 IoT 서비스 Slides 0 2022

PROJECT.1

석사 연구 1

REGEN: RECURRENT ENSEMBLE
METHODS FOR GENERATIVE MODELS

01

ABOUT PROJECT

저의 주 연구 분야인 생성 모델에서 양상블에 대한 논문입니다. 석사 졸업 논문이고 현재 2023 IEEE Access에 투고 준비 중입니다.

REGEN: Recurrent Ensemble Methods for Generative Models

Ahjeong Park, Youngmi Park, Chulyun Kim

양상블은 여러 모델을 활용하여 단일 구성 모델보다 더 나은 예측 성능을 얻습니다. 대부분의 기존 양상블은 모델이 블랙박스로 간주되어 최종 결과만을 취합합니다. 이러한 고려는 다양한 종류의 기계학습 모델에 양상을 모델을 적용할 수 있게 합니다. 특히 최종 출력의 Diversity가 제한적인 Discriminative 모델은 모델 간 의견 수렴이 쉽기 때문에 기존의 양상을 적용하기에 적합했습니다. 하지만 Generative 계열의 모델은 최종 출력의 길이와 범위에 제한이 없고 Diversity가 높아 모델 간 합의를 보는데 문제가 있습니다. 따라서 이 문제를 고려하기 위해 Generative 계열의 모델에 대한 **새로운 양상블인 RGEN을 제안했습니다.**

새로운 양상블은 **Consensus, Survival Ensemble**입니다. 또한 기존의 양상블을 새롭게 재해석한 Majority Ensemble도 설계하여 비교 실험을 진행했습니다.

양상을 구성 모델로 Seq2Seq, Transformer을 활용했고 각 모델의 Decoder는 매 단계마다 합의를 진행한 후 다음 생성에 영향을 미치며 양상을 진행합니다. 기계번역 및 문자열 사칙연산에 대해 실험 결과, **REGEN은 단일 구성 모델 뿐만 아니라 기존 양상을 보다 성능이 우수함을 확인했습니다.**

기여도

실험 설계  95

실험 진행  100

논문 작업  95

현재 상태

2023년 3월 IEEE Access 투고 준비 중입니다.

총 3가지 양상블 방법 제안

- 1) Baseline: Majority
- 2) Recurrent Ensemble with Survival
- 3) Recurrent Ensemble with Consensus

실험 방법

각 양상블에 대해 2개의 Case Study를 진행했습니다.

Case Study: Application of Recurrent Ensemble

- 1) Seq2Seq
- 2) Transformer

REGEN: Recurrent Ensemble Methods for Generative Models

Ahjeong Park, Youngmi Park, Chulyun Kim

1-1) Majority Ensemble in Seq2Seq

- Generative(RNN) 구조에 전통적인 양상을 방법을 재해석한 Baseline 양상을 방법입니다.
- 전통적인 양상을과 동일하게 time-step의 중간 출력을 고려하지 않고 각 모델의 최종 output을 결합합니다.

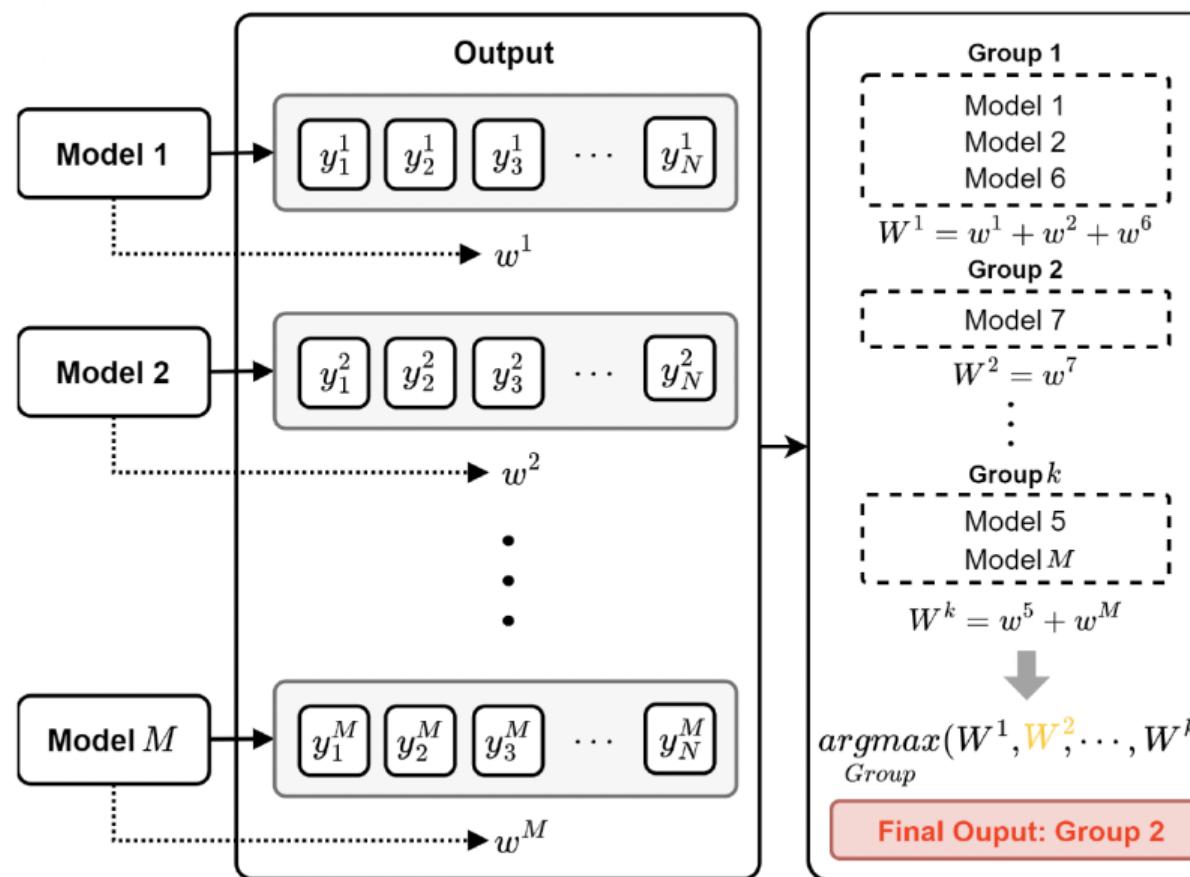


Figure 3.1: Overall Structure of Traditional Majority in Seq2Seq

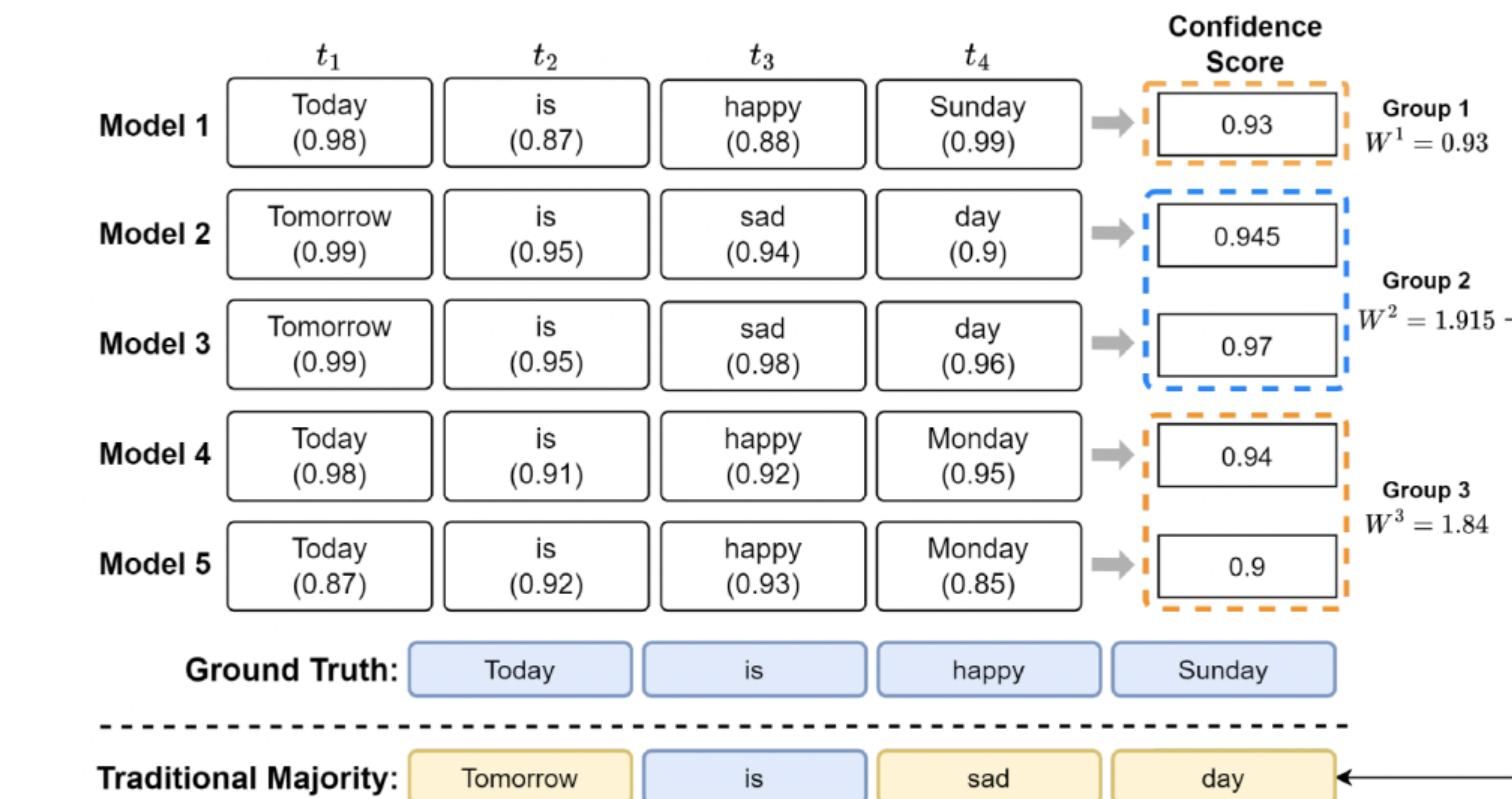


Figure 3.2: A toy example: Traditional Majority

REGEN: Recurrent Ensemble Methods for Generative Models

Ahjeong Park, Youngmi Park, Chulyun Kim

1-2) Survival Ensemble in Seq2Seq

- Recurrent Ensemble의 한 종류로, 게임 방식과 비슷해 Survival로 명칭했습니다.
- 살아남은 Winner(승자) 모델만이 다음 time-step 예측에 참여할 수 있도록 해서 최종 결과를 결정합니다.

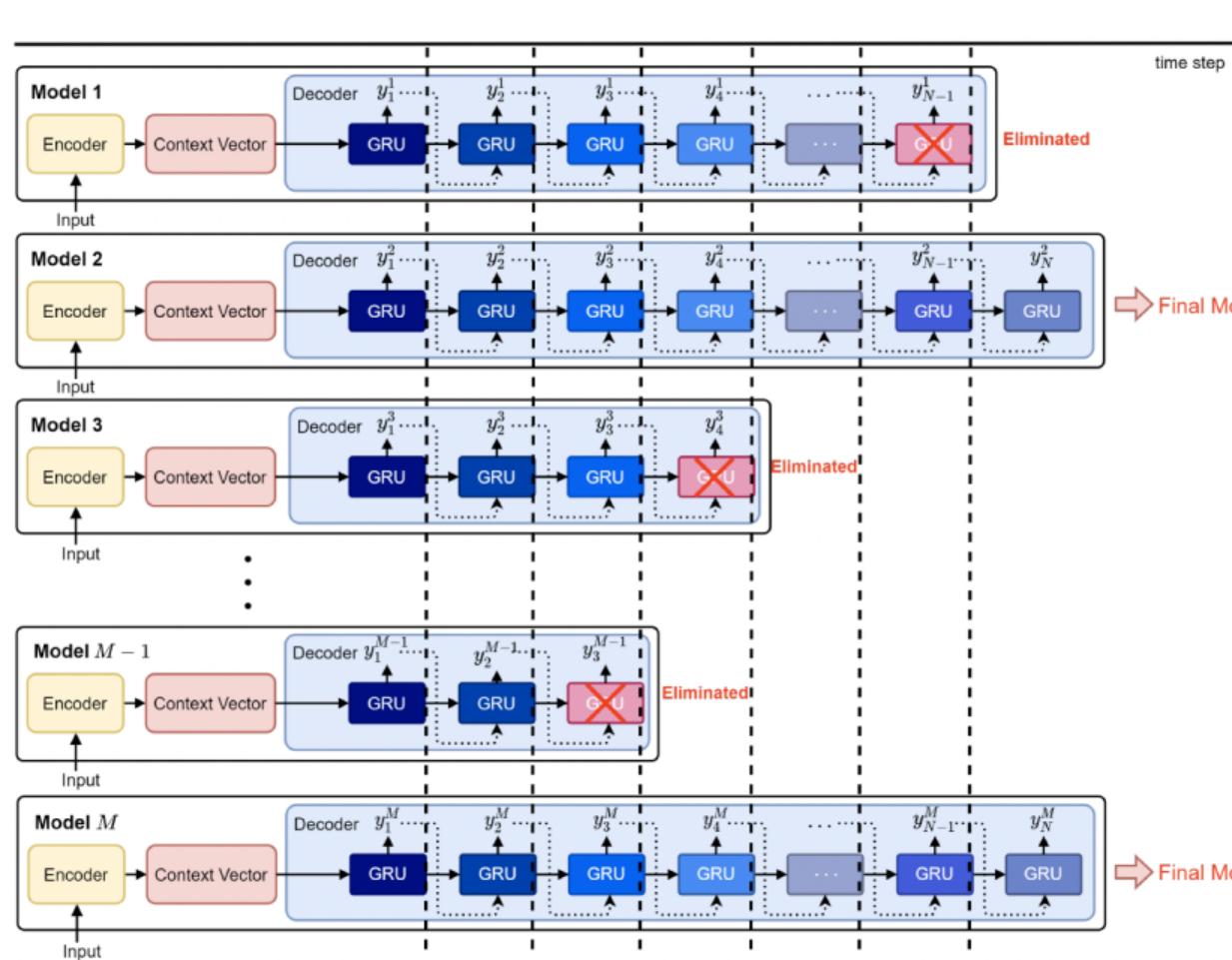


Figure 3.3: Overall Structure of Survival in Seq2Seq

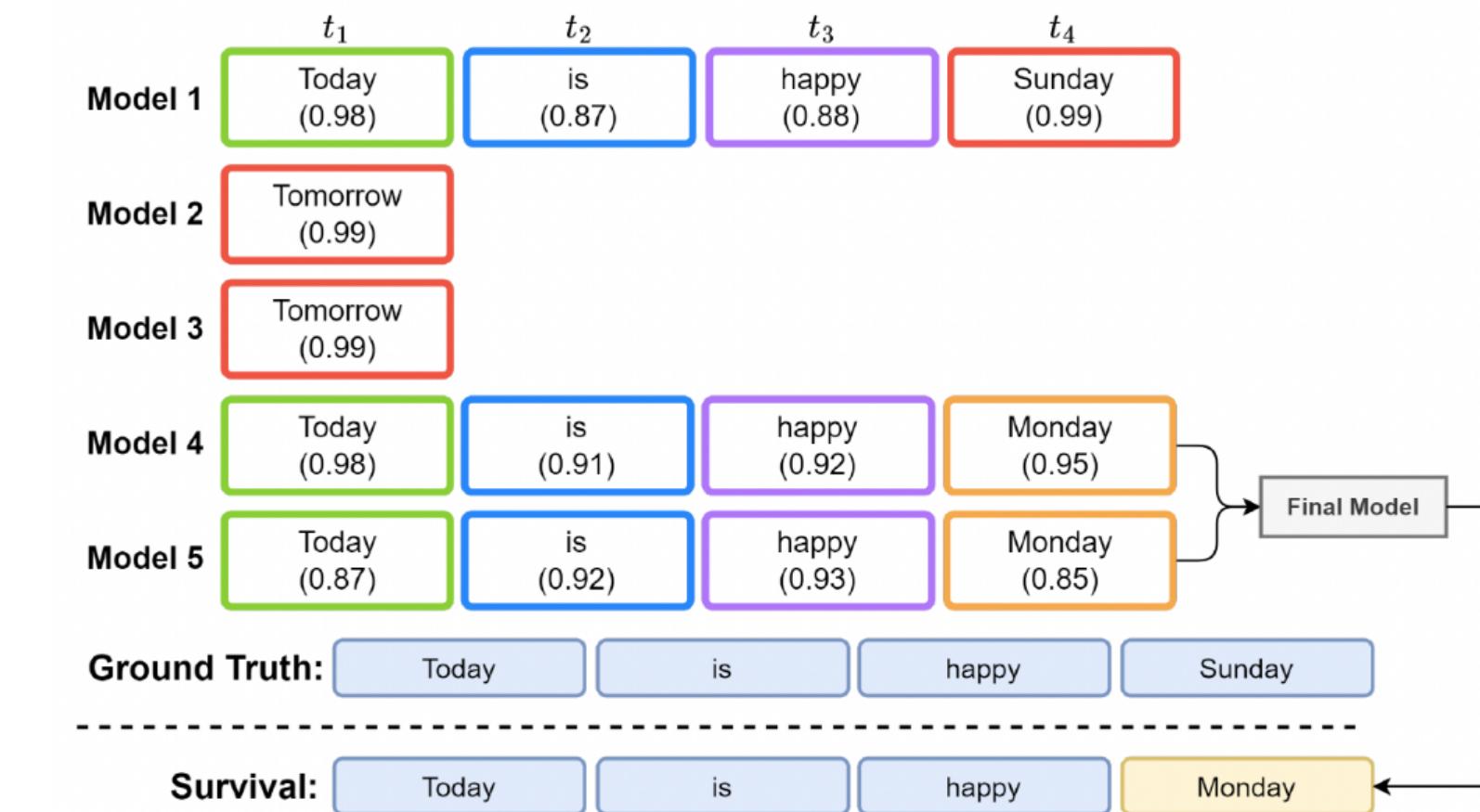


Figure 3.4: A toy example: Recurrent Ensemble using Survival

REGEN: Recurrent Ensemble Methods for Generative Models

Ahjeong Park, Youngmi Park, Chulyun Kim

1-3) Consensus Ensemble in Seq2Seq

- Recurrent Ensemble의 한 종류로, 가장 높은 성능을 달성한 양상을 방법입니다.
- 모든 모델의 매 Time-step의 중간 출력을 고려합니다.
- 단일 모델의 Confidence를 고려한 Voting 결과를 넘겨주기 위해 Hard Voting이 아닌 Soft Voting 방식을 선택했습니다.

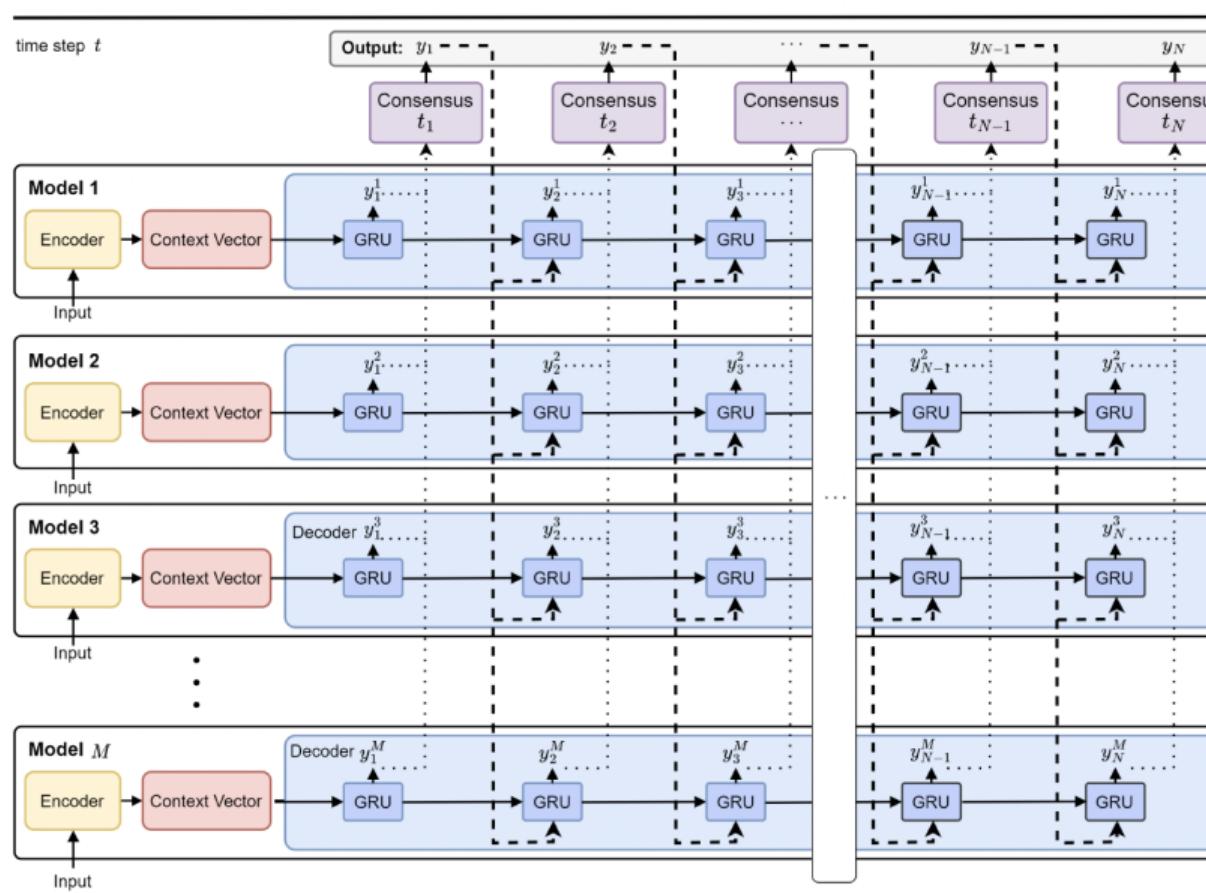


Figure 3.5: Overall Structure of Consensus in Seq2Seq

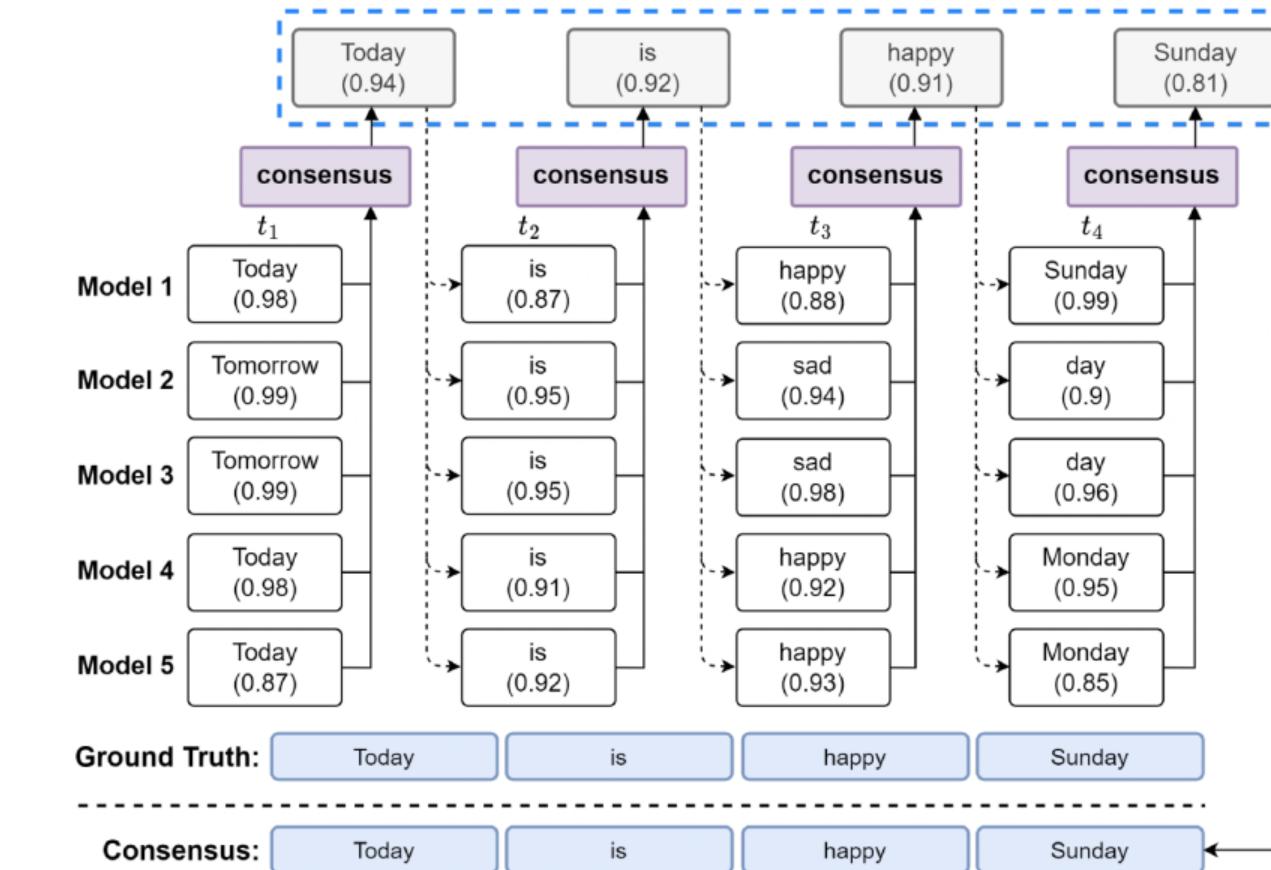


Figure 3.6: A toy example: Recurrent Ensemble using Consensus

REGEN: Recurrent Ensemble Methods for Generative Models

Ahjeong Park, Youngmi Park, Chulyun Kim

2) Majority, Survival and Consensus Ensemble in Transformer

- Transformer 모델에서도 양상을 알고리즘은 동일하게 적용됩니다.
- 한가지 다른 점은 Transformer는 Seq2Seq과 달리 이전의 output이 하나씩 되먹임 되는 방식이 아니라 마스킹 되는 부분을 고려해주었습니다.

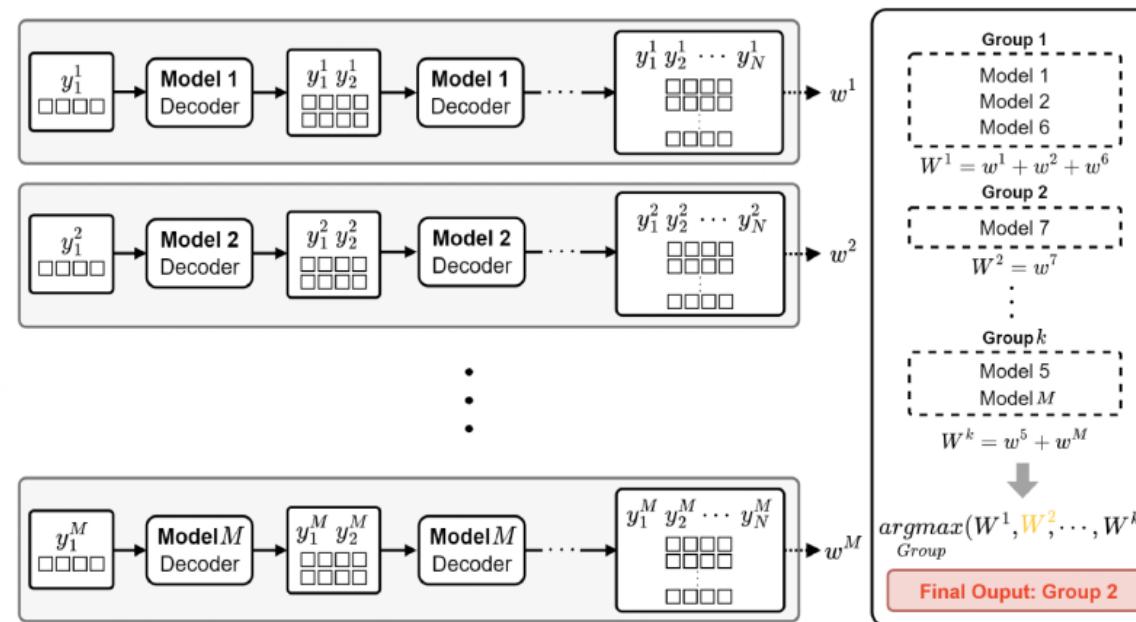


Figure 3.7: Overall Structure of Traditional Majority in Transformer

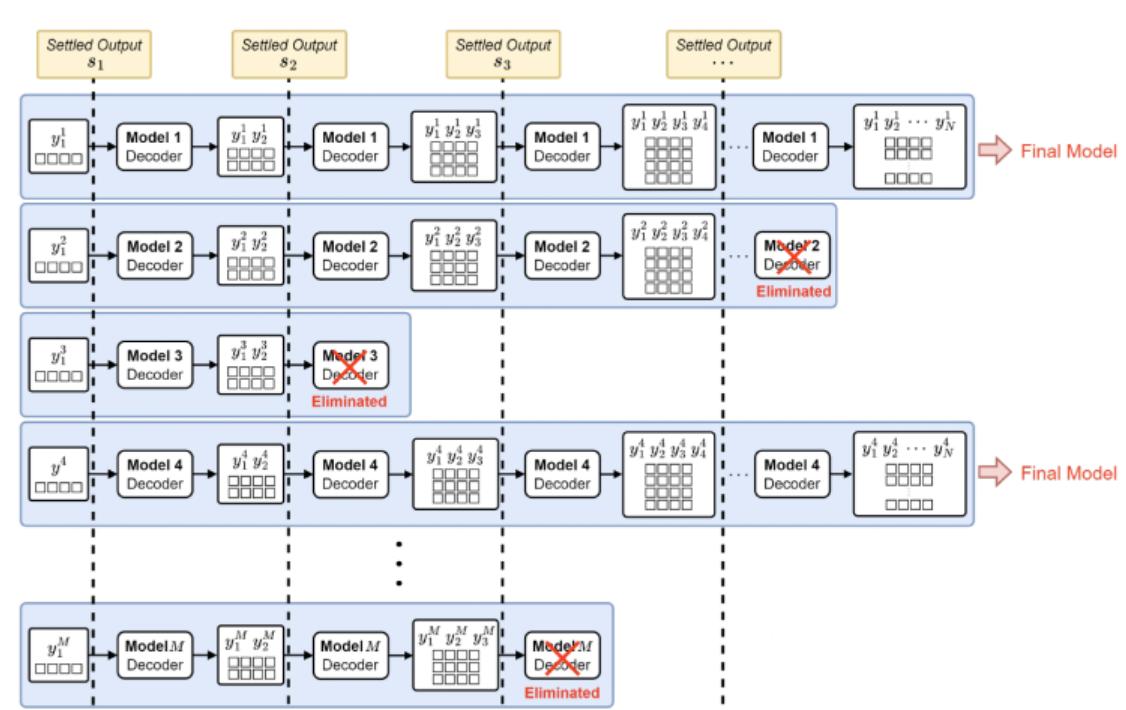


Figure 3.9: Overall Structure of Survival in Transformer

Majority Ensemble

Survival Ensemble

Consensus Ensemble

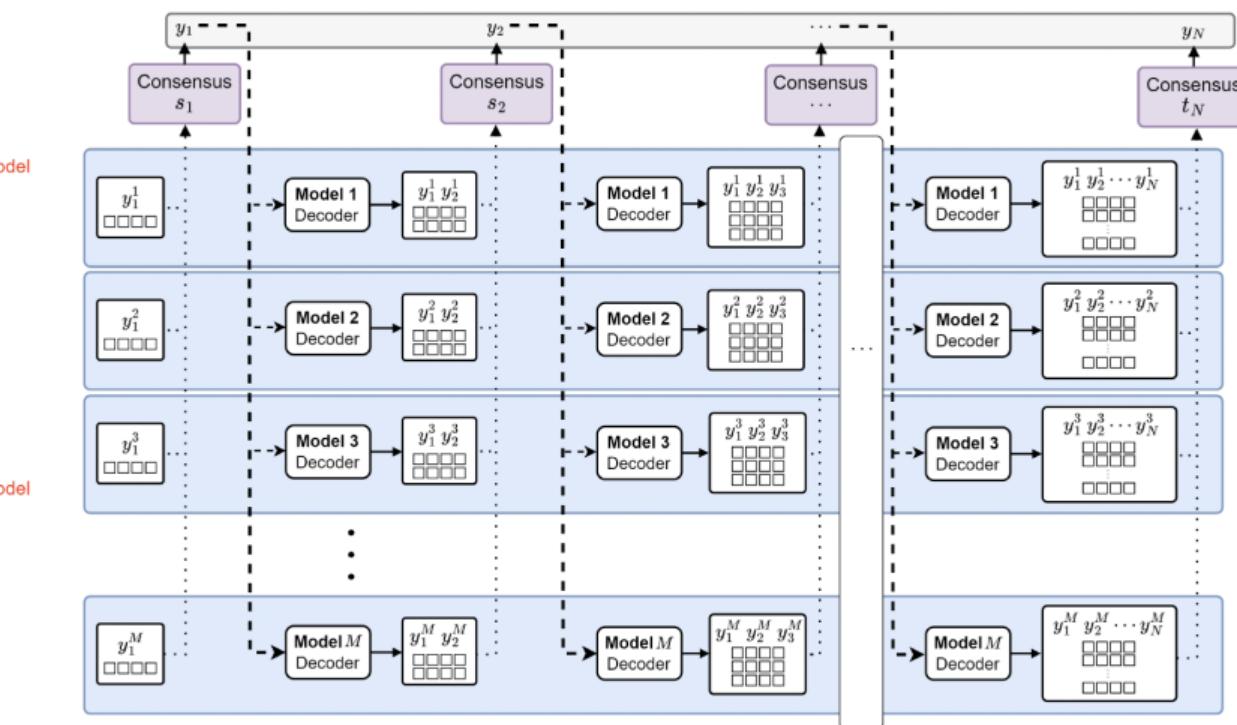


Figure 3.11: Overall Structure of Consensus in Transformer

REGEN: Recurrent Ensemble Methods for Generative Models

Ahjeong Park, Youngmi Park, Chulyun Kim

Experiments Setup

Seq2Seq

- 1) Neural Machine Translation (Spain-English)
 - 단일 모델 총 15개
 - Baselines: Majority, Independent Ensemble
 - Metric: TQE(Translation Quality Estimation, BERTScore, BLEU, ROUGE)
- 2) String Arithmetic
 - 단일 모델 총 5개
 - Metric: Accuracy

Transformer

- 1) Neural Machine Translation(German-English)
 - 단일 모델 총 10개
 - Baselines: Majority, Checkpoint Ensemble
 - Metric: TQE(Translation Quality Estimation, BERTScore, BLEU, ROUGE)

REGEN: Recurrent Ensemble Methods for Generative Models

Ahjeong Park, Youngmi Park, Chulyun Kim

Experiments Results

Neural Machine Translation in Seq2Seq

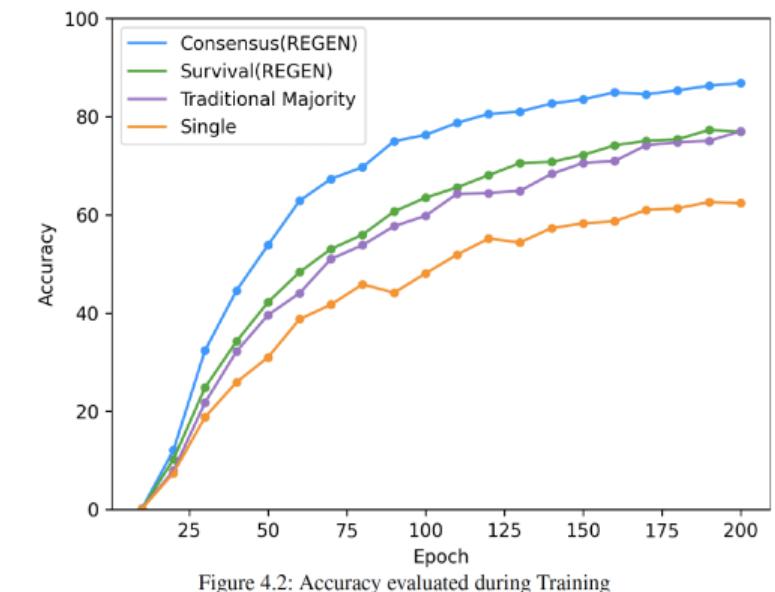
Table 4.2: Comparison of the performance of TQE (%), BLEU (%), and F1 BERT (%) for various ensemble methods and single model (avg) in Spanish-English machine translation

Model	TQE	BLEU	F1 BERT
Independent (top 1) [35]	60.34	19.49	78.40
Independent (top 2) [35]	62.59	20.5	78.91
Independent (top 3) [35]	62.83	20.64	79.06
Traditional Majority	72.43	21.9	80.20
Survival (REGEN)	73.10	21.28	80.31
Consensus (REGEN)	74.33	22.71	80.53
Single (Avg, 10 epoch)	69.09	19.56	79.28

String Arithmetic in Seq2Seq

Table 4.7: Comparison of performance of Accuracy for various ensemble methods and single model in String Arithmetic

Model	Accuracy (%)
Traditional Majority	76.38
Survival (REGEN)	77.53
Consensus (REGEN)	86.75
Single (Avg)	64.70
Single 1	65.46
Single 2	68.32
Single 3	67.90
Single 4	58.14
Single 5	63.70



- 본 연구의 3가지 양상을 모두 단일 모델을 뛰어 넘은 것을 확인할 수 있었습니다.
- Recurrent Ensemble > 기존 양상들(Independent Ensemble)
- Recurrent Ensemble > baseline: Majority Ensemble
- Consensus > Survival > Majority
- 기계 번역 실험과 마찬가지로 여전히 Recurrent Ensemble이 기존의 양상을 방법 보다 뛰어남을 확인할 수 있었습니다.

REGEN: Recurrent Ensemble Methods for Generative Models

Ahjeong Park, Youngmi Park, Chulyun Kim

Experiments Results

Neural Machine Translation in Transformer

Table 4.9: Comparison of the performance of TQE (%), BLEU (%), and F1 BERT (%) scores for the various ensemble methods and single model in German-English machine translation

Model	TQE	BLEU	F1 BERT
Checkpoint (190, 195, best) [9]	81.23	23.62	92.91
Checkpoint (190, 195) [9]	81.5	23.76	93.01
Traditional Majority	83.10	24.02	93.56
Survival (REGEN)	84.48	25.04	94.21
Consensus (REGEN)	84.55	25.15	94.23
Single (Avg, 200 Epoch)	84.19	24.93	94.09

Table 4.10: Comparison of the performance of n-gram BLEU (%) for various ensemble methods and single model (avg) in

Model	BLEU 1	BLEU 2	BLEU 3	BLEU 4
Checkpoint (190, 195, best) [9]	41.68	32.27	21.08	13.07
Checkpoint (190, 195) [9]	41.76	32.43	21.21	13.2
Traditional Majority	42.3	32.6	21.43	13.38
Survival (REGEN)	42.63	33.96	22.55	14.14
Consensus (REGEN)	42.8	34.09	22.65	14.21
Single (Avg, 200 Epoch)	42.7	33.78	22.41	14.05

Table 4.8: TQE, BLEU, F1 BERT 성능 비교

- Recurrent Ensemble > Baseline(Majority, 기존 양상블)
- Recurrent Ensemble > Single Models

Table 4.9: N-gram BLEU 성능 비교

- Recurrent Ensemble > Majority > Checkpoint Ensemble(기존 양상블)
- Recurrent Ensemble > Single Model

PROJECT.2

02

석사 연구 2

ALSI-TRANSFORMER: TRANSFORMER-BASED CODE
COMMENT GENERATION WITH ALIGNED LEXICAL AND
SYNTACTIC INFORMATION

2023 IEEE Access

ABOUT PROJECT

정부 과제에 참여하면서 진행한 연구입니다.

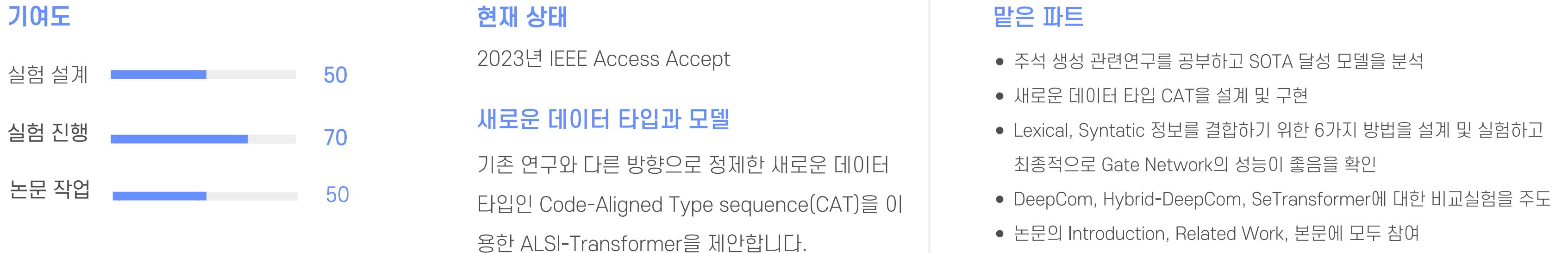
ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information

Youngmi Park, Ahjeong Park, Chulyun Kim

소프트웨어 프로젝트의 복잡성과 업데이트 빈도가 증가하면서 프로그램 이해의 중요성이 증가합니다. 이 때 좋은 주석은 프로그램 이해의 효율성 증가에 결정적인 역할을 합니다. 하지만 직접 주석을 작성하는 것은 시간이 많이 걸리고 품질을 보장하기 어려우며 기존 코드 주석은 관련 코드의 발전에 따라 계속해서 업데이트 돼야 합니다. **따라서 자동으로 고품질 주석을 생성하는 것은 매우 중요합니다.**

본 논문에서는 자동 주석 생성의 정확도를 향상시키기 위해 다음을 소개합니다. **Lexical 정보와 Syntactic 정보의 순서와 길이를 정렬하기 위한 새로운 구문 시퀀스인 CAT(Code-Aligned Type sequence)을 제안하고 그에 따른 신경망 모델인 ALSI-Transformer을 제안합니다.** ALSI-Transformer는 Transformer 기반 딥러닝 모델로, 함수 단위의 소스코드를 입력으로 넣었을 때 적절한 자연언어 주석을 출력으로 생성하는 것을 목표로 합니다. 특히, CAT과 Gate Network를 활용해 소스코드의 lexical, syntactic 정보를 결합합니다.

다양한 실험을 통해, 표준 기계 번역 메트릭을 사용하여 논문의 방법을 현재 Baselines와 비교했고 **이 방법이 코드 주석 생성에서 최첨단 성능**을 달성한다는 것을 보였습니다.



ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information

Youngmi Park, Ahjeong Park, Chulyun Kim

1) CAT(Code-Aligned Type sequence)

- 딥러닝 기반 자동 주석 생성 모델에 쓰이는 소스코드 구조 정보 데이터인 Abstract Syntax Tree(AST), Structure-Based Traversal(SBT) 등을 수집 후 기존 연구에서 활용하는 소스코드 데이터의 한계점을 분석하고 이를 보완하는 **새로운 데이터 타입**을 설계했습니다.
- 기존 연구와 다른 방향으로 정제한 새로운 데이터 타입인 Code-Aligned Type sequence(CAT)을 제안하고 이를 활용해 **소스코드의 의미적, 구조적 정보 모두 추출**하는 방법을 설계했습니다.
- CAT은 소스코드 정보 손실과 중복 및 불필요한 정보 포함 등의 기존 연구의 한계점을 보완하여 **정보 손실을 최소화**합니다.
- CAT은 소스코드에서 생성된 AST로부터 **코드 토큰과 타입 토큰이 순서에 따라 정렬되어 추출된 정보로 기존의 SBT 정보보다 더 나은 syntactic 정보**를 얻을 수 있습니다.

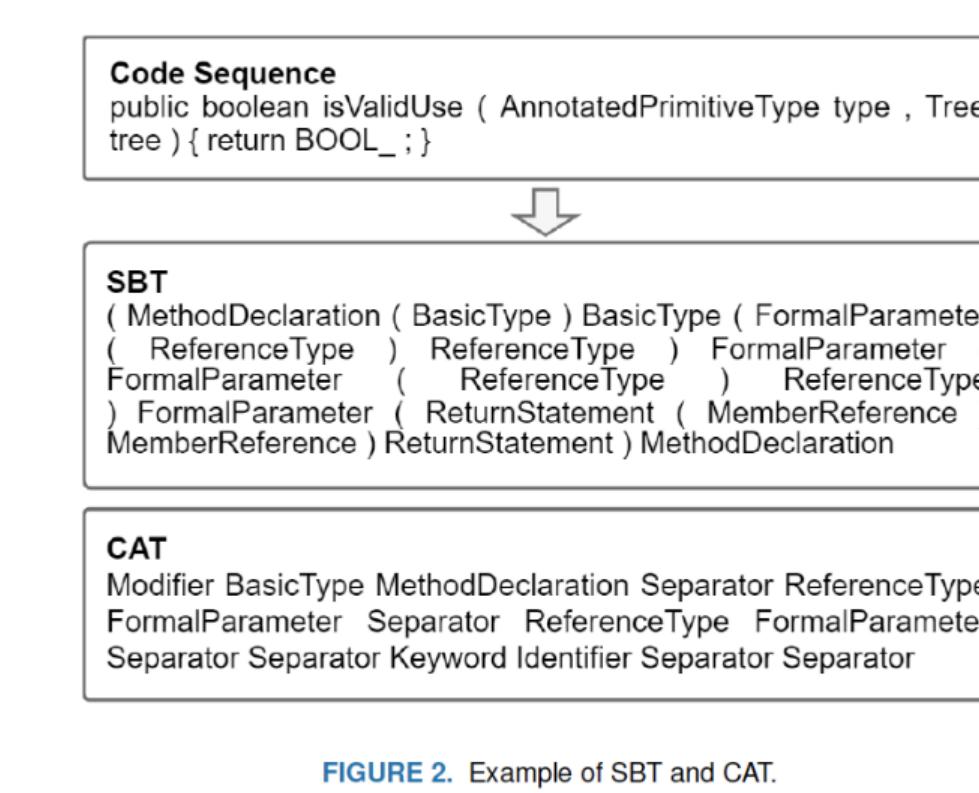


FIGURE 2. Example of SBT and CAT.

ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information

Youngmi Park, Ahjeong Park, Chulyun Kim

2) ALSI-Transformer

- 소스코드 파일의 하위 단위인 **함수에 대해 자연언어 설명 생성 알고리즘**을 구현해 새로운 모델인 **ALSI-Transformer**을 제안
- ALSI-Transformer는 Transformer 기반 딥러닝 모델로 함수 단위의 소스코드를 입력으로 넣었을 때 적절한 자연언어 주석을 출력으로 생성하는 것을 목표로 함. 특히, **CAT과 Gate Network**를 활용해 소스코드의 **lexical, syntactic 정보를 결합**했습니다.

TABLE 8. Results of comparison between ALSI-Transformer (*Gate Network*) and five aggregation methods in terms of BLEU and METEOR.

Aggregation Method	BLEU	METEOR
<i>Alternation</i>	51.26	64.05
<i>Separation</i>	52.40	65.20
<i>Addition</i>	53.21	65.89
<i>Average</i>	50.75	63.46
<i>Concatenation</i>	51.41	64.23
<i>Gate Network</i>	53.80	66.11

- Lexical, Syntactic 2개의 정보를 결합하기 위한 1개의 인코더를 사용했습니다.
- 2개의 정보를 결합하는 방법으로 Alternation, Separation, Addition, Average, Concatenation, Gate Network를 실험했고 **최종 결과 Gate Network의 방법이 높은 성능**을 보였습니다.
- 따라서 ALSI-Transformer는 Gate Network를 사용합니다.
- 자세한 결합 방법은 다음장에 준비되어 있습니다.

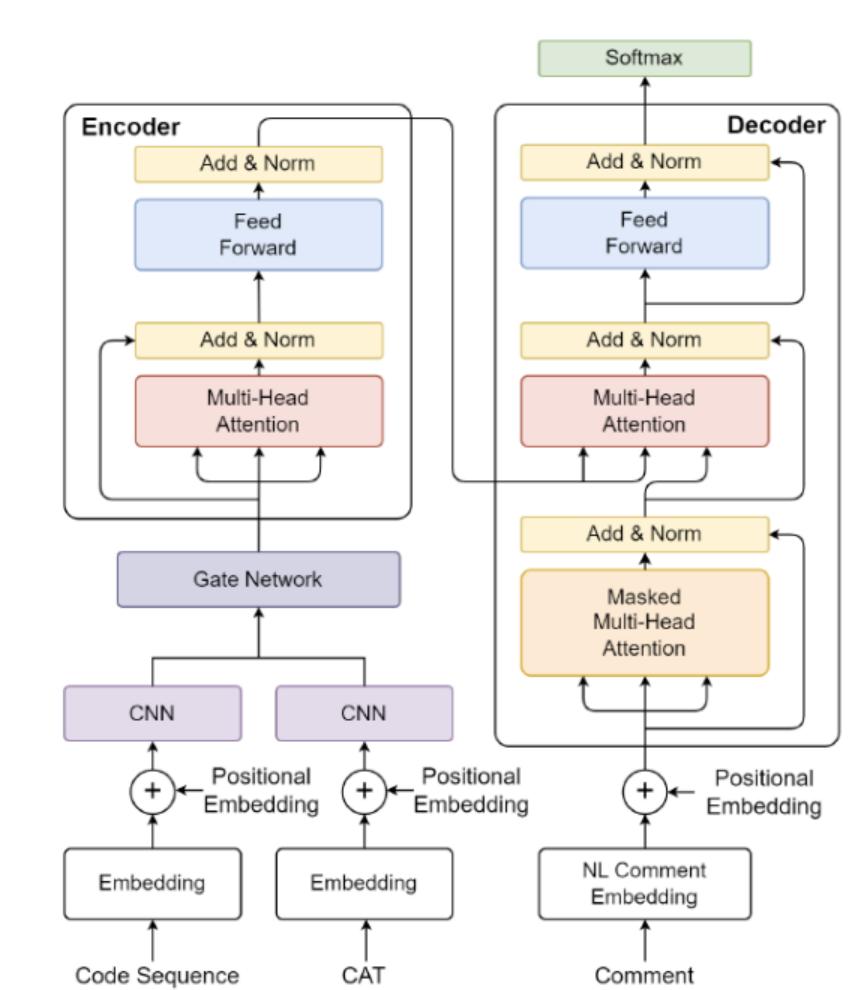


FIGURE 1. Structure of the ALSI-Transformer.

ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information

Youngmi Park, Ahjeong Park, Chulyun Kim

2-1) ALSI-Transformer - Aggregation Methods

- 1) Alternation: 코드 시퀀스와 CAT이 번갈아 나오는 결합 방법입니다.
- 2) Separation: 코드 시퀀스 묶음과 CAT 묶음이 '<code>'라는 토큰으로 연결되는 결합 방법입니다.
- 3) Addition: 코드 시퀀스 임베딩과 CAT 임베딩을 Add 한 결합 방법입니다.
- 4) Average: 코드 시퀀스 임베딩과 CAT 임베딩을 Average 한 결합 방법입니다.
- 5) Concatenation: 코드 시퀀스와 임베딩과 CAT 임베딩을 Concatenation한 결합 방법입니다.

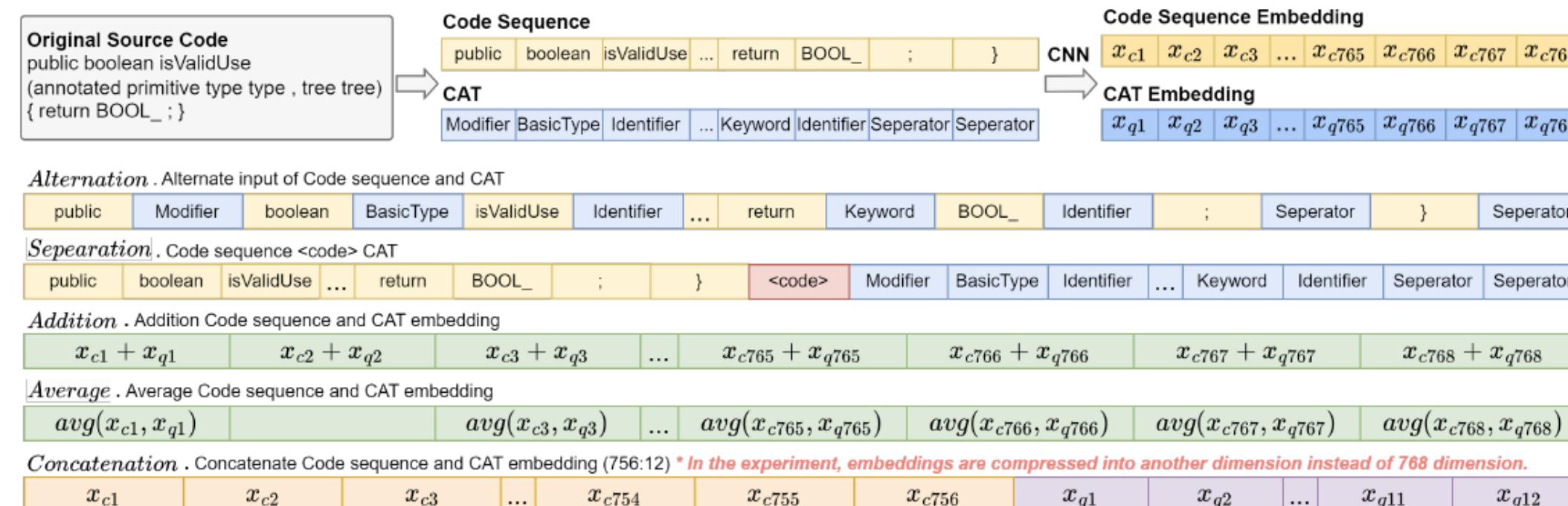


FIGURE 3. Example of five aggregation methods. Code sequence and CAT extracted from original source code used for *Alternation* and *Separation* methods. Code sequence embedding and CAT embedding are generated through the CNN layer. Two embeddings are used for *Addition* and *Average*. *Concatenation* makes embedding in different ratios.

ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information

Youngmi Park, Ahjeong Park, Chulyun Kim

Experiment Results

- 6개의 베이스라인과 비교하기 위해 BLEU, N-BLEU, METEOR에 대해 성능 평가를 진행했습니다.
- 논문의 방법이 주석 생성에서 SOTA(2022년 7월 기준)을 달성함을 확인했고 이 방법이 Lexical 과 Syntactic 정보 결합이 중요하다는 것을 입증했습니다.
- 기존의 방법보다 모델 사이즈, 파라미터 개수가 적고 학습시간이 효율적임을 확인했습니다.
- 또한 2개의 정보를 처리할 때 two encoder와 one encoder 중 one encoder의 성능이 좋음을 확인했습니다.

TABLE 5. BLEU, n-gram BLEU, and METEOR score for our model ALSI-Transformer compared with six baselines.

Models	BLEU	BLEU 1	BLEU 2	BLEU 3	BLEU 4	METEOR
Seq2Seq (attention) [9]	37.87	46.53	41.53	37.81	35.04	23.29
Transformer [10]	45.55	55.62	46.30	41.57	38.69	29.06
DeepCom [6]	20.26	32.88	21.91	18.93	17.35	31.72
Hybrid-DeepCom [11]	38.20	51.63	40.56	36.70	34.41	51.26
ComFormer [7]	42.99	55.31	47.57	41.72	36.26	59.12
SeTransformer [4]	49.00	62.78	51.91	47.47	44.62	62.99
ALSI-Transformer	53.80	57.26	56.29	52.49	50.03	66.11

TABLE 7. Comparison of the number of model parameters, training time, and model size between ALSI-Transformer and SeTransformer.

Models	# of Parameters	Training Time(s)	Model Size (GB)
ALSI-Transformer	146,939,910	174,279	8.3
SeTransformer [10]	167,308,038	265,140	10

TABLE 9. Comparison results between ALSI-Transformer and ALSI-Transformer (two-encoder).

Models	BLEU	METEOR	Model Size (GB)
ALSI-Transformer	53.80	66.11	8.3
ALSI-Transformer (two-encoder)	50.05	63.30	9.5

감사합니다!

잘 부탁드립니다!

Potfolio

CONTACT

ahjeong@sookmyung.ac.kr
010 7448 8798

