

AI_ERC

All notion logs are available at <https://star-tray-8b0.notion.site/4a410a2b1eea449dabc197506caa4b43?pvs=4>

| 각자 역할

김도훈

DataLoader 완성

- Video 및 Audio가 없는 데이터들이 존재
- 데이터 전처리 (결측치에 대한 처리)

모델 실험

- AST 모델 실험
 - Waveform → Mel Spectrogram 코드 구현, 데이터 전처리도 수행함
- Swin3D/B 실험
- Linear classifier 실험 결과 정리
 - 각 feature에 대해 라벨로의 분류
 - Video, Audio에 비해 Text가 과하게 높은 현상에 대한 고민

DF-ERC 논문 리뷰, DDM 및 CFM 파이프라인 구축

- 너무 무거운 모델이 만들어지므로 기각
- DDM에서 사용된 Feature Disentanglement Method를 차용, 이를 구현하고 현 데이터에 적용

모델 구성

- 2가지 모델 (vanilla, feature fusion) 구축

보고서

- LATEX 보고서 양식 생성, overleaf에서 공유할 수 있도록 사용
- .tex 문법에 대한 이해 및 자료 작성을 위한 기초 문법 안내

발표자료

- PPT 템플릿 바탕으로 제작
- table of contents 바탕으로 순서를 결정하고, 이에 대한 역할 분배

박나연

Audio Data Feature Extraction

- audio data 처리 방식 공부 : video → audio → mel spectrogram
- AST, Omnivec, UAVM 등 audio feature extractor 논문 리뷰 & ast 모델 선택
- AST model 구현 & extract audio feature

Vanilla Neural Network

- 코드 실행 & 가장 성능이 높게 나오도록 hyperparameter 조정
- accuracy & loss plot

Disentanglement 논문 코드 구현

- CRM 구현해보려 했으나 실패

보고서

- 작성

발표자료

- 제작

박세용

Text DataLoader 모델에 맞게 보완 및 완성

- DeBERTaV3 모델에 맞게 text data 전처리

모델 실험

- DeBERTaV3 모델 실험
 - Text data를 feature extraction 하는 코드 구현
 - 모델의 훈련과 검증 정확도 및 손실 계산
 - 위의 결과에 대한 confusion matrix 수행 BERT, DeBERTa, DeBERTaV3 논문 리뷰
 - MELD dataset 자체가 크기가 크지 않기 때문에 이를 효과적으로 처리할 수 있는 모델에 대해 고민함.
 - DeBERTa 모델의 한계 파악, 'tug-of-war' 현상을 해결할 수 있는 모델인 DeBERTaV3를 찾음. 보고서

- Feature extraction 영역에서 가장 가중치가 높았던 text data set에 대한 추가적인 실험 진행.
- 데이터 class의 불균형을 해결하기 위한 방안으로 weighted class method와 oversampling을 해결 방안을 제시 및 이에 대한 코드 구현 및 실행 발표자료
- Video, Audio, Text 각각의 데이터들에 대한 feature extraction을 진행한 모델을 발표

미팅 날짜

5/8 Meeting,"May 8, 2024",
 5/15 Meeting,"May 15, 2024",
 5/22 Meeting,"May 22, 2024",
 5/29 Meeting,"May 29, 2024",
 6/2 Meeting,"June 2, 2024",
 6/5 Meeting,"June 5, 2024",
 6/8 Meeting,"June 8, 2024",

각자 리뷰한 논문들

Title	Who	Paper Link
Revisiting Disentanglement and Fusion on Modality and Context in Conversational Multimodal Emotion Recognition	김도훈	https://arxiv.org/pdf/2308.04502
A Transformer-based Model with Self-distillation for Multimodal Emotion Recognition in Conversations	김도훈	https://arxiv.org/pdf/2310.20494v1
Video Swin Transformer	김도훈	https://arxiv.org/pdf/2106.13230

DEBERTA: DECODING- ENHANCED BERT WITH DIS ENTANGLED ATTENTION	박세 용	https://arxiv.org/abs/2006.03654
DEBERTAV3: IMPROVING DEBERTA USING ELECTRA-STYLE PRE-TRAINING WITH GRADIENT DISENTANGLED EMBEDDING SHARING	박세 용	https://arxiv.org/abs/2111.09543
ANGLE- OPTIMIZED TEXT EMBEDDINGS	박세 용	https://arxiv.org/abs/2309.12871
Sample Design Engineering: An Empirical Study of What Makes Good Downstream Fine-Tuning Samples for LLMs	박세 용	https://arxiv.org/abs/2404.13033
AST: Audio Spectrogram Transformer	박나 연	https://arxiv.org/abs/2104.01778
OmniVec: Learning robust representations with cross modal sharing	박나 연	https://arxiv.org/abs/2311.05709
A Transformer- based Model with Self-distillation for Multimodal Emotion	박나 연	https://ieeexplore.ieee.org/document/10109845

Recognition in Conversations

VATT:

Transformers for
Multimodal Self-
Supervised
Learning from
Raw Video, Audio
and Text

박나
연

<https://arxiv.org/abs/2104.11178>

BERT: Pre-
training of Deep
Bidirectional
Transformers for
Language
Understanding

박세
용

<https://arxiv.org/abs/1810.04805>

Attention Is All
You Need

박세
용

<https://arxiv.org/abs/1706.03762>

ELECTRA: PRE-
TRAINING TEXT
ENCODERS AS
DISCRIMINATORS
RATHER THAN
GENERATORS

박세
용

<https://arxiv.org/abs/2003.10555>