

Prompt 기반의 Full-Shot Learning과 Few-Shot Learning을 이용한 알츠하이머병 치매와 조현병 진단

정민교^o, 나승훈,
김고운, 신병수, 정영철
Jungmk@jbnu.ac.kr^o



Jeonbuk National University*
Jeonbuk National University Hospital

목 차

- **서론**

- 치매 및 조현병 배경지식
- Prompt-tuning 적용 배경

- **Prompt 기반의 Full-Shot Learning과 Few-Shot Learning을 이용한 알츠하이머병 치매와 조현병 진단**

- Prompt-tuning approach
- Input Data Structure
- 분류 과정

- **실험**

- Task Definition
- Dataset
- 실험 Setting
- 실험 결과

- **결론 및 향후 연구**

- 결론
- 향후 연구

서론

- 치매 - 배경지식

: 기억, 언어, 판단력 등 여러 영역의 인지기능이 감소해 일상 생활에 지장을 초래할 수 있는 질병

의사: 자, 아버님이 오늘 집에서 여기까지 오시는데 어떻게 오셨는지 오시는 길에 보이셨던 건물이나 풍경들 기억나는 것들 말씀해주세요.

환자: 예, / 예, 예, 저 아 아 앞에 그 저 아 앞 오 앞오 밤 시 식사를 허고 예 저 그냥 그냥 응 여 까지 헤아꼬, 아직 여끼지 거시기 없이...

의사: 자, 아버님 여기까지 오시는데 교통 수단 뭘 타고 오셨어요?

환자: 저 거시기 말하자면은 같이 어 우리 같이 한 동네 뭐 이 같이 있는 음 그 누구죠?

- ① “그”, “저”와 같은 대명사의 잦은 빈도
- ② 질문에 대한 답변 지연 -> 5초 이상 답변이 지연됨을 의미하는 태그 ‘I’
- ③ 질문과 관련 없는 발언

서론

- 조현병 – 배경지식

: 잦은 망상과 와해된 언어 사용의 증상을 보이는 중증 정신 질환

의사: 지금까지 ‘왜 살지’라고 생각하면서 부정적인 생각이 들었던 적 있으세요?

환자_1: 그죠, 만약에 핸드폰도 막 아 이거 전화 올 사람도 없지 하면서 이것도 이 핸드폰만 쓰구 다른 거 같지 않고 그러고 죽었으면 좋겠다. 그런 생각이 들었어.

의사: 최근 기분이 처지거나 우울한 적이 있으십니까?

환자_2: 예, 누가 내 발을 자꾸 건드는 거 같아 갖고 밤에 와가지고.

- ① 우울한 느낌의 단어 사용
- ② 망상과 관련된 발언

서론

- Prompt-tuning 적용 배경

- # Prompt-tuning method

- ① PTR: Prompt Tuning with Rules for Text Classification [Han et al '21]

- : prompt Tuning with rules를 통해 many-class text classification 태스크 수행

- ➡ Prompt-tuning method를 통한 Classification task 성능 향상

- # Fine-tuning method

- ① Dementia detection using transformer-based deep learning and natural language processing models [P. Saltz et al '21]

- : BERT, ALBERT, XLNet 등 트랜스포머에 기반한 모델을 fine-tuning 하여 치매 분류 태스크 수행

- ② Assessing schizophrenia patients through linguistic and acoustic features Using Deep Learning Techniques [Huang et al '22]

- : BERT를 fine-tuning 하여 Text의 Semantic Textual Feature를 추출하고, 조현병 분류 태스크 수행

- ➡ Prompt-Tuning method를 적용한 연구 부족

서론

- Prompt-tuning 적용 배경

확보한 Dataset 특징

- ① Language Models are Few-Shot Learners [Brown et al '20]
- ② PPT: Pre-trained Prompt Tuning for Few-shot Learning [GU et al '22]

➡ 적은 데이터를 활용해 학습하는 Few-Shot learning을 통해 성능 향상

- Prompt-tuning approach

 - # Template 구성

 - : 입력 문장에 대해 설명하는 임의의 문장을 구성해 추가

 - # [MASK] 토큰 예측

 - 구성한 템플릿 내에 [MASK] 토큰을 삽입해 예측

‘명량’ 영화는 재미있어.

< Input Sentence >

이 영화는 [MASK]

< Template >



➡ [MASK] token에 대한 MLM Head 학습

➡ CLS 토큰을 이용하는 일반적인 Fine-tuning method와 비교 비교해 언어모델에 친화적

Prompt 기반의 Full-Shot Learning과 Few-Shot Learning을 이용한 알츠하이머병 치매와 조현병 진단

- Input Data structure

① 일상 생활에 대해 묻는 질문, MMSE 치매 선별 검사에 쓰이는 질문과 같이 특정 프로토콜에 따라 내담자와 대화한 데이터

② $subject = [s_1^{doctor}, s_1^{client}, \dots, s_N^{doctor}, s_N^{client}]$

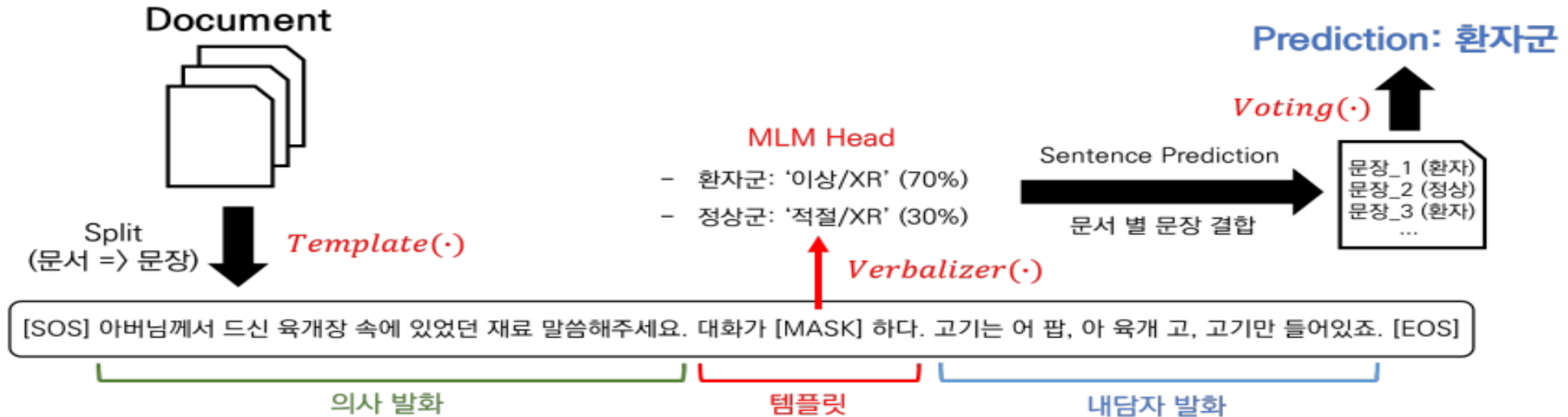
: $subject$ 는 분류 대상 하나를 의미하는 단위로서, $|s_i^{doctor}|$ 발화 길이를 가진 의사 발화 $|s_i^{doctor}|$ 와 $|s_i^{client}|$ 발화 길이를 가진 내담자 발화 s_i^{client} N 개로 구성

③ $qa_i = [s_i^{doctor}, T, s_i^{client}]$, $d_i \in V^{\max_len}$

: qa_i 는 i 번째 의사 발화와 템플릿, 내담자 발화를 의미, d_i 는 qa_i 를 모델에 입력하기 위해 token id로 변환한 것으로 padding token id를 추가해 최대 입력 사이즈 \max_len 으로 통일, V 는 모델에서 사용한 vocab의 사이즈를 의미

Prompt 기반의 Full-Shot Learning과 Few-Shot Learning을 이용한 알츠하이머병 치매와 조현병 진단

• 분류 과정



① '대화가 [MASK] 하다.' 템플릿을 사용해 Input data 구성

- 환자군을 의미하는 '이상/XR' 토큰
- 정상군을 의미하는 '적절/XR' 토큰

② Evaluation

- 문서별로 문장을 다시 구성해 voting 방법으로 문서에 대한 예측 값 추론

실 험

- Task Definition

치매 관련 Task

① D_AD (알츠하이머성 치매 환자군 vs 정상군)

② D_MCI (경도인지장애 환자군 vs 정상군)

: 치매 전 단계로 알려져 있는 경도인지장애 환자군과 정상군 분류 태스크

③ D_AMY (아밀로이드 양성 환자군 vs 아밀로이드 음성 환자군)

: 치매의 원인이라는 의학적 소견이 있는 독성 단백질 아밀로이드에 대해 양성을 보이는 치매 환자군과 음성을 보이는 치매 환자군 분류

실 험

- Task Definition

조현병 관련 Task

① S_SZ (조현병 환자군 vs 정상군)

② S_Spectrum (조현병 스펙트럼 환자군 vs 정상군)

: 조현병을 포함하는 기타정신병을 보이는 환자군과 정상군을 분류하는 태스크

③ S_Global (인지기능 결함 환자군_전역적 vs 결함 없는 환자군)

: 인지기능 영역에서 결함을 보이는 조현병 환자군과 결함을 보이지 않는 환자군을 분류하는 태스크

③ S_Local (인지기능 결함 환자군_2개 vs 결함 없는 환자군)

: 2가지 인지기능 영역에서 결함을 보이는 조현병 환자군과 결함을 보이지 않는 환자군을 분류하는 태스크

실험

- Dataset

치매 Task

Task	학습 데이터	검증 데이터	평가 데이터
D_AD	38 : 161	5 : 20	5 : 20
D_MCI	116 : 161	15 : 20	15 : 20
D_AMY	30 : 38	4 : 6	4 : 6

조현병 Task

Task	학습 데이터	검증 데이터	평가 데이터
S_SZ	159 : 125	20 : 16	20 : 16
S_Spectrum	190 : 125	24 : 16	24 : 16
S_Global	60 : 92	7 : 12	8 : 12
S_Local	85 : 71	11 : 9	11 : 9

: 학습 데이터 (8) : 검증 데이터 (1) : 평가 데이터 (1)

- 실험 Setting

- macro/micro F1 score 기준 평가 (매 epoch 마다 macro, micro 별 Best score 모델 저장)
- 3개의 seed 값을 사용해 평균 성능을 제시
- 확보한 데이터셋 모두를 사용하는 Full-shot learning과 적은 데이터셋을 사용하는 Few-shot learning 사용

실험

- 실험 결과

Task	Full-Shot (macro/micro F1)		Few-Shot (macro/micro F1)	
	Baseline	Prompt Tuning	Baseline	Prompt Tuning
D_AD	75.08/88.00	75.08/88.00	68.60/84.00	44.44/80.00
D_MCI	64.83/72.86	68.84/72.86	54.34/60.00	38.32/57.14
D_AMY	41.28/60.00	43.50/60.00	41.28/60.00	48.75/60.00
S_SZ	70.67/75.00	73.92/79.17	39.28/51.39	53.17/54.17
S_Spectrum	73.68/76.25	70.91/71.25	38.20/53.75	53.49/58.75
S_Global	36.49/60.00	52.26/60.00	62.20/60.00	57.76/ 60.00
S_Local	44.36/47.50	43.66/45.00	46.68/47.50	46.53/55.00

표 4. Full-Shot 및 Few-Shot 실험 결과

- 빨간 폰트는 macro/micro F1 score 모두 성능 향상, 파란 폰트는 하나의 F1-score만 성능 향상
- Full-shot learning에서는 macro- F1 score 기준 4개 태스크 성능 향상, micro F1 score 1개 태스크 성능 향상
- Few-shot learning에서는 macro F1 score 기준 3개 태스크 성능 향상, micro F1 score 기준 3개 태스크 성능 향상

결론 및 향후 연구

- **결론**

: 일반적인 도메인을 사전 학습한 언어 모델의 내부 지식을 병원 분야와 같은 특수한 도메인에 활용하여도 성능이 향상될 수 있다는 결과를 도출하였다.

- **향후 연구**

- ① 적절한 템플릿 및 레이블 구성에 대한 추가 연구

- : Auto prompting 관련 연구 추가 진행

- ② 부족한 데이터에 문제 해결 관련 추가 연구

- : 데이터 증강 관련 연구 추가 진행

Thank You