



개별 연구 9주차 회의록

회의내용	성능 평가 및 논문 초안 작성		
교과목명	개별연구	담당교수	신연순
과제명	재난 및 위기관리 성능평가 연구	팀원	김관우, 서혜민
회의록	1. 학습한 네가지 모델의 성능 평가		
	1) 성능평가는 f1 score 을 사용하여 진행(Kaggle 에서 제공하는 public score 가 f1 score 를 의미함)		
	2) .네 가지 모델은 정규화 방법에 따라 나뉘지며 각각 아래와 같음		
	- 정규화 적용 안함(none)		
	submission(none).csv 9 minutes ago by KWANWOO KIM add submission details		0.83144
	- dropout 적용(dropout)		
	submission(dropout).csv 9 minutes ago by KWANWOO KIM add submission details		0.83420
	- dropconnect 적용(dropconnect)		
	submission(dropconnect).csv just now by KWANWOO KIM add submission details		0.83450
	- dropout + maxnorm 적용(maxnorm)		
	submission(maxnorm).csv 21 hours ago by KWANWOO KIM add submission details		0.83818
	2. 논문 초안 작성		
	- 전자 정보 통신 학술대회 참가를 위해 양식에 맞는 논문 작성		
	- 참가비와 관련해서 지원금 등의 조사 필요		
	3. 추후 계획		
	- 논문 수정 및 학술 대회 참가		

개별 연구 8주차 회의록

회의내용	BERT 모델 학습(dropout + maxnorm) 및 batch size 조사		
교과목명	개별연구	담당교수	신연순
과제명	재난 및 위기관리 성능평가 연구	팀원	김관우, 서혜민
회의록	<p>4. Batch size 및 max-norm 조사</p> <p>3) Batch size 는 전체 데이터 셋을 여러 그룹으로 나눠서 학습을 시킬 때, 하나의 그룹에 해당하는 데이터 수를 의미. Batch size 는 클수록 학습 횟수가 줄어들어 속도가 향상되지만, overfitting 이 일어날 수 있고, 작을수록 학습 횟수가 늘어나기 때문에 가중치의 잦은 업데이트로 불안정해질 수 있음.</p> <p>4) Max-norm 은 신경망의 가중치를 크게 설정되지 못하게 제한하는 방법으로 주로 과적합을 피해 모델의 매개변수를 제한. Dropout 과 maxnorm 은 서로 상호보완적이기 때문에, 같이 사용했을 때 효과가 매우 좋음.</p> <p>5. 기존 dropout 에 maxnorm 를 적용하여 학습 수행</p> <pre>dense_layer1 = Dense(units=256, activation='relu', kernel_constraint=max_norm(2.))(clf_output) dense_layer1 = Dropout(0.5)(dense_layer1) out = Dense(1, activation='sigmoid')(dense_layer1)</pre> <p>기존 학습과 마찬가지로 [학습 계수: (epochs = 8, batch_size = 10)]로 설정. max_norm 은 일반적으로 2~4 를 사용하므로, 2 로 설정.</p> <p>6. 팜 경진대회 신청</p> <p>해당 개별 연구 주제를 바탕으로 경진대회를 신청.</p> <p>7. 추후 계획</p> <ul style="list-style-type: none"> - 논문 작성 - 학습 결과 분석 		

개별 연구 7주차 회의록

회의내용	BERT 모델 학습(nodrop, dropout, dropconnect)		
교과목명	개별연구	담당교수	신연순
과제명	재난 및 위기관리 성능평가 연구	팀원	김관우, 서혜민
회의록	<p>8. dropout 과 dropconnect 추가 학습</p> <p>9. 참고 논문과 같은 조건의 학습이 불가능하다는 문제가 발생하여 해결 방법 도출</p> <p>1) nodrop, dropout, dropconnect 를 각각 학습</p> <p>2) 학습한 세가지 모델에 대해 성능을 비교</p> <p>3) TV 의 뉴스와 같은 매체의 데이터보다 sns 의 재난 관련데이터가 실시간으로 많은 정보를 얻을 수 있어서 위기 관리에 사용되면 좋은 효과를 가져올 수 있을 것으로 예상되어 연구를 진행하였고, BERT 를 이용하면 성능이 좋은 NLP 기술로 재난 트윗 분류가 가능할 것으로 예상 여기서 BERT 를 사용하여 재난 관련 트윗을 분류할 때 정규화를 nodrop, dropout, dropconnect 세가지를 같이 적용해 보고 어떤 정규화를 진행했을 때 가장 성능이 좋은 재난 트윗 분류 모델을 학습할 수 있는지 비교하는 과정으로 논문을 작성할 예정</p> <p>10. nodrop, dropout, dropconnect 를 각각 학습수행</p> <p>[dropout]</p> <pre>dense_layer1 = Dense(units=256, activation='relu')(clf_output) dense_layer1 = Dropout(0.5)(dense_layer1) out = Dense(1, activation='sigmoid')(dense_layer1)</pre> <p>[dropconnet]</p> <pre>dense_layer1 = Dense(units=256, activation='relu')(clf_output) dense_layer1 = DropConnectDense(units=256, prob=0.5, activation='relu', use_bias=True)(dense_layer1) out = Dense(1, activation='sigmoid')(dense_layer1)</pre> <p>[학습 계수: (epochs = 8, bach_size = 10)]</p> <p>11. 추후 계획</p> <ul style="list-style-type: none"> - 학습한 세 가지 모델의 성능 비교 - 논문 초안 작성 		

개별 연구 6주차 회의록

회의내용	BERT 모델 input data 및 기존 논문과의 차별성 조사		
교과목명	개별연구	담당교수	신연순
	과제명	재난 및 위기관리 성능평가 연구	팀원 김관우, 서혜민
회의록	1. input 데이터 분석		
	<pre>['[CLS]', 'our', 'deeds', 'are', 'the', 'reason', 'of', 'this', 'earthquake', 'may', 'allah', 'forgive', 'us', 'all', '[SEP]'] [101, 2256, 15616, 2024, 1996, 3114, 1997, 2023, 8372, 2089, 16455, 9641, 2149, 2035, 102, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,</pre>		
	<p>- 세번째 줄의 pad_mask 는 attention_mask 라고 하며, token 의 길이인 15 만쯤 실제로 데이터가 채워져 있기 때문에 True 를 의미하는 1, 그 뒤의 패딩은 데이터가 채워져 있지 않기 때문에 False 를 의미하는 0 을 가지며, 이 pad_masks 를 통해 불필요한 attention 은 건너뛰게 된다.</p>		
	<p>- 네번째 줄의 segment_id 는 token_type_ids 로 입력으로 들어오는 text 를 양 끝에 CLS 와 ELS 토큰이 붙는 하나의 단일 문장으로 들어오기 때문에 0 으로 설정되었으며, 그 뒤의 패딩은 어떠한 값을 가져도 의미가 없으므로 0 으로 설정.</p>		
	2. 기존 논문과의 차별성 조사		
	<p>- 기존 논문은 0.5 dropout layer 및 relu 함수가 적용된 dense layer 추가</p> <p>We added a dropout layer of 0.5 to the BERT layer's output to complete the pre-trained model, and then we said a Dense layer with relu activation. We use the relu and sigmoid activation functions. The architecture will generate opportunities for each class, real disaster (target) or non-real disaster (non-target). Figure 7 shows a summary of the model.</p> <p>- 차별점</p> <p>1) droupout layer 를 두 번 적용</p> <p>2) droupout layer 대신 dropconnect layer 적용.</p>		

개별 연구 5주차 회의록

회의내용

BERT 모델 encoding 구조 및 input data 분석

교과목명

개별연구

담당교수

신연순

과제명

재난 및 위기관리 성능평가 연구

팀원

김관우, 서혜민

회의록

1. encoding 구조 학습

- attention 벡터를 생성하는 함수의 실행 과정 및 input 학습
- attention 들을 여러 개를 모아 multi head attention 을 생성하고 이는 토큰 간의 더욱 복잡한 관계를 파악
- 인코더 구조의 feed forward 레이어는 현재 인코더에서 다음 인코더와 연결하기 위해 출력 벡터의 크기를 인풋으로 들어온 입력 벡터의 크기와 같게 함
- BERT 의 pretraining 단계에서는 Masked Language Model 과 Next Sentence Prediction 을 사용해 학습을 진행하고 이를 통해 문장의 표현을 학습
- 위와 같은 단계를 거쳐 pretrain 된 모델에 class label 의 개수만큼 output 을 가지는 Dense Layer 를 붙여서 fine-tuning 을 수행

2. 인코딩 데이터 분석

- 지난주에 pretraining 까지 직접 코드를 작성해 수행했다고 생각했지만 이번 주에 BERT 구조를 학습한 결과 아직 pretraining 은 진행하지 않았고, 데이터 토큰화와 인코딩 단계 까지만 수행된 것으로 파악
- BERT 모델에 사용할 train 데이터 분석
- token -> 각 단어를 토큰화 하여 int 형으로 매핑 됨

[101, 2256, 15616, 2024, 1996, 3114, 1997, 2023, 8372, 2089, 16455, 9641, 2149, 2035, 102,

- mask_pad -> 토큰들은 전부 1 로 토큰이 아닌 패딩은 0 으로 설정됨
- [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
- Segment_id -> 토큰화 된 토큰이 어떤 문장에 속하는지 0 또는 1 을 통해 지정하는데 단일 문장은 모두 0 으로 지정한다.

[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

3. 추후 계획

- 토큰화와 인코딩 된 데이터를 사용하여 BERT 모델의 Pre-Training 수행

개별 연구 4주차 회의록

회의내용

BERT 모델 pre-train 및 모델 분석

교과목명

개별연구

담당교수

신연순

과제명

재난 및 위기관리 성능평가 연구

팀원

김관우, 서혜민

1. Pre-train

- tensorflow에서 BERT 모델을 불러오고, 이전 단계에서 특수기호, 문장부호 및 소문자 처리의 data cleaning 을 완료한 data 에서 모델의 입력으로 들어갈 token, mask, segment 를 추출하기 위해 모델의 FullTokenizer 를 불러온다.

```
module_url = "https://tfhub.dev/tensorflow/bert_en_uncased_L-24_H-1024_A-16/1"
bert_layer = hub.KerasLayer(module_url, trainable=True)

vocab_file = bert_layer.resolved_object.vocab_file.asset_path.numpy()
do_lower_case = bert_layer.resolved_object.do_lower_case.numpy()
tokenizer = tokenization.FullTokenizer(vocab_file, do_lower_case)
```

- bert_encode 함수 내에서 불러온 FullTokenizer 를 통해 text 를 token, mask, segment 로 변환한다.

```
text = tokenizer.tokenize(text)
text = text[:max_len-2]
input_sequence = ["[CLS]"] + text + ["[SEP]"]
pad_len = max_len - len(input_sequence)

tokens = tokenizer.convert_tokens_to_ids(input_sequence)
tokens += [0] * pad_len
pad_masks = [1] * len(input_sequence) + [0] * pad_len
segment_ids = [0] * max_len

train_input = bert_encode(train.text.values, tokenizer, max_len=160)
test_input = bert_encode(test.text.values, tokenizer, max_len=160)
train_labels = train.target.values
```

2. BERT 모델 및 데이터 분석.

- BERT 모델의 트랜스포머 인코더 내부 과정 및 data 전처리 필요성 분석.
- BERT 활용 코드 분석(<https://www.kaggle.com/friskycodeur/nlp-with-disaster-tweets-bert-explained>)

3. 추후 계획

- BERT 모델 분석 및 train Model.

회의록

개별 연구 3주차 회의록

회의내용

데이터 전 처리 및 모델 분석

교과목명

개별연구

담당교수

신연순

과제명

재난 및 위기관리 성능평가 연구

팀원

김관우, 서혜민

1. 데이터 전처리

- Pandas 와 preprocess_kgptalkie 를 사용하여 데이터 전 처리

3913	5561	flood	boston	when fizzy is sitting in the regular flood seats and no one kn	0
3914	5563	flood	United Sta	JKL cancels Flash Flood Warning for Bell Harlan Knox [KY] http://	1
3915	5564	flood	New York	Spot Flood Combo 53inch 300W Curved Cree LED Work Light B	1
3916	5565	flood	New York	Spot Flood Combo 53inch 300W Curved Cree LED Work Light B	0
3917	5567	flood	Warrantyt	A [small] flood with big??consequences https://t.co/CVPdVHxd1F	1
3918	5569	flood	British Col	Homecoming Queen Killed on Way Home from the Prom by Flo	1
3919	5570	flood		2pcs 18W CREE Led Work Light Offroad Lamp Car Truck Boat N	0

[이전 데이터 -> 특수문자와 대문자 포함]

5561	flood	boston	when fizzy is sitting in the regular flood seats and no
5563	flood	United Sta	jkl cancels flash flood warning for bell harlan knox ky
5564	flood	New York	spot flood combo 53inch 300w curved cree led work
5565	flood	New York	spot flood combo 53inch 300w curved cree led work
5567	flood	Warrantyt	a small flood with bigconsequences
5569	flood	British Col	homecoming queen killed on way home from the pr
5570	flood		2pcs 18w cree led work light offroad lamp car truck

[이후 데이터 -> 특수문자 삭제 및 대문자를 소문자로 변환]

2. 논문의 BERT 모델과 용어 분석

- 기존 모델의 연구 방식, BERT 모델의 레이어, 성능 평가 방법 등 분석
- BERT 활용 코드 분석(<https://www.kaggle.com/chandrakanthlns/disaster-classification-tweets>)

회의록

개별 연구 2주차 회의록

회의내용

연구 계획 재수립

교과목명

개별연구

담당교수

신연순

과제명

재난 및 위기관리 성능평가 연구

팀원

김관우, 서혜민

3. 주제 선정

- 논문 Disaster Tweets Classification in Disaster Response using Bidirectional Encoder Representations from Transformer(BERT) 을 참고하여 주제 선정
- BERT 모델 학습 방법 - <https://www.kaggle.com/c/nlp-getting-started#> 내 disaster dataset 와 BERT 오픈 소스 코드를 참고하여 BERT 모델 학습.
- BERT 모델 검증 방법 - 학습된 BERT 모델에 트위터에서 크롤링을 통해 수집한 실제 데이터를 적용하여 테스트.
- BERT 모델 학습을 통한 홍수 재난 트윗 분류를 주제로 선정
- BERT 모델의 테스트는 2021 년 3 월 22 일에 호주에서 발생한 홍수를 기준으로 수행.

4. 주제 선정 이유

- 양방향으로 학습이 되는 BERT 모델을 통해 정교하며, 높은 정확도의 자연어 처리(NLP)를 달성
- 학습한 모델을 실제 데이터에 적용하여 BERT 모델의 성능 확인 가능.

5. 프로젝트 일정 조율

재난 및 위기관리 성능평가 연구 (트위터 데이터를 활용한 재난 연관 트윗 분류)

재난 및 위기관리 성능평가 연구 (트위터 데이터를 활용한 재난 연관 트윗 분류)								동국대학교 컴퓨터공학과 - 김관우, 서혜민			
	8/30~										~11/14
	1주차	2주차	3주차	4주차	5주차	6주차	7주차	8주차	9주차	10주차	11주차
관련 논문 분석											
세부 계획 수립											
데이터 수집 / 전처리											
인공지능 모델 학습											
데이터 크롤링											
성능 평가 및 테스트											
논문 작성											

회의록

개별 연구 1주차 회의록

회의내용	연구 계획 수립																																																																																																																																												
교과목명	개별연구					담당교수	신연순																																																																																																																																						
과제명	재난 및 위기관리 성능평가 연구					팀원	김관우, 서혜민																																																																																																																																						
회의록	6. 주제 선정																																																																																																																																												
	<ul style="list-style-type: none">- 논문 An Automated Early Alert System for Natural Disaster Risk Reduction: A Review 과 A Cloud Based Disaster Management System 을 참고하여 주제 선정- 재난 상황 인식 방법 - 첫 번째 논문의 SNS 데이터를 이용한 홍수 감지 시스템 채택- 재난 상황 전달 방법 - 두 번째 논문의 3D 모델링 방법 채택- 트위터 크롤링을 통한 홍수 상황 인식 및 3D 모델 개발을 주제로 선정- 연구는 2021 년 3 월 22 일에 호주에서 발생한 홍수를 기준으로 수행																																																																																																																																												
	7. 주제 선정 이유																																																																																																																																												
	<ul style="list-style-type: none">- TV 나 라디오보다 SNS 를 사용하면 더 많은 정보를 수집 및 사용 가능- 3D 시뮬레이션을 웹으로 제공하면 도시에서 울리는 경보 이외에 3D 로 시각화 된 홍수 재난 상황을 인식 가능																																																																																																																																												
	8. 프로젝트 일정 조율																																																																																																																																												
	<table><tr><td colspan="11">재난 및 위기관리 성능평가 연구 (sns 데이터를 사용한 홍수 감지 시스템)</td></tr><tr><td></td><td colspan="10">동국대학교 컴퓨터공학과 - 김관우, 서혜민</td></tr><tr><td></td><td>8/30~</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>~11/14</td></tr><tr><td></td><td>1주차</td><td>2주차</td><td>3주차</td><td>4주차</td><td>5주차</td><td>6주차</td><td>7주차</td><td>8주차</td><td>9주차</td><td>10주차</td><td>11주차</td></tr><tr><td>관련 논문 분석</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>세부 계획 수립</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>데이터 수집 / 전처리</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>인공지능 모델 학습</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>웹 시뮬레이션 제작</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>성능 평가 및 테스트</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>논문 작성</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table>											재난 및 위기관리 성능평가 연구 (sns 데이터를 사용한 홍수 감지 시스템)												동국대학교 컴퓨터공학과 - 김관우, 서혜민											8/30~										~11/14		1주차	2주차	3주차	4주차	5주차	6주차	7주차	8주차	9주차	10주차	11주차	관련 논문 분석												세부 계획 수립												데이터 수집 / 전처리												인공지능 모델 학습												웹 시뮬레이션 제작												성능 평가 및 테스트												논문 작성											
	재난 및 위기관리 성능평가 연구 (sns 데이터를 사용한 홍수 감지 시스템)																																																																																																																																												
		동국대학교 컴퓨터공학과 - 김관우, 서혜민																																																																																																																																											
		8/30~										~11/14																																																																																																																																	
		1주차	2주차	3주차	4주차	5주차	6주차	7주차	8주차	9주차	10주차	11주차																																																																																																																																	
관련 논문 분석																																																																																																																																													
세부 계획 수립																																																																																																																																													
데이터 수집 / 전처리																																																																																																																																													
인공지능 모델 학습																																																																																																																																													
웹 시뮬레이션 제작																																																																																																																																													
성능 평가 및 테스트																																																																																																																																													
논문 작성																																																																																																																																													