ERNIE 2.0: A CONTINUAL PRE-TRAINING FRAMEWORK FOR LANGUAGE UNDERSTANDING

발표자 : 이기창
https://ratsgo.github.io

ERNIE2.0

- Motivation

  대부분의 pretrained model이 co-occurrence로만 학습
  named entity, sentence order, semantic similarity 등도 중요

ERNIE2.0

- Motivation

대부분의 pretrained model이 co-occurrence로만 학습
named entity, sentence order, semantic similarity 등도 중요

➔ Knowledge Integration (ERNIE)

➔ Continual Learning (ERNIE2.0)

ERNIE2.0

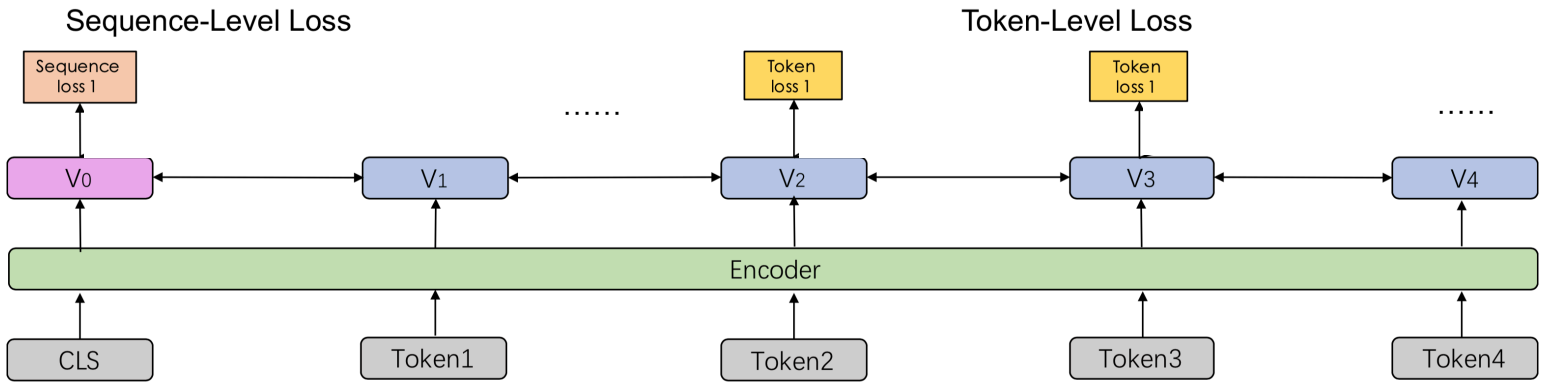
- Continual Learning

  여러 task를 순차적으로 학습
  단 새 task를 학습할 때 이전 task도 같이 학습해
  이전 학습 내용을 잊어버리지 않도록 함

ERNIE2.0

- Continual Learning

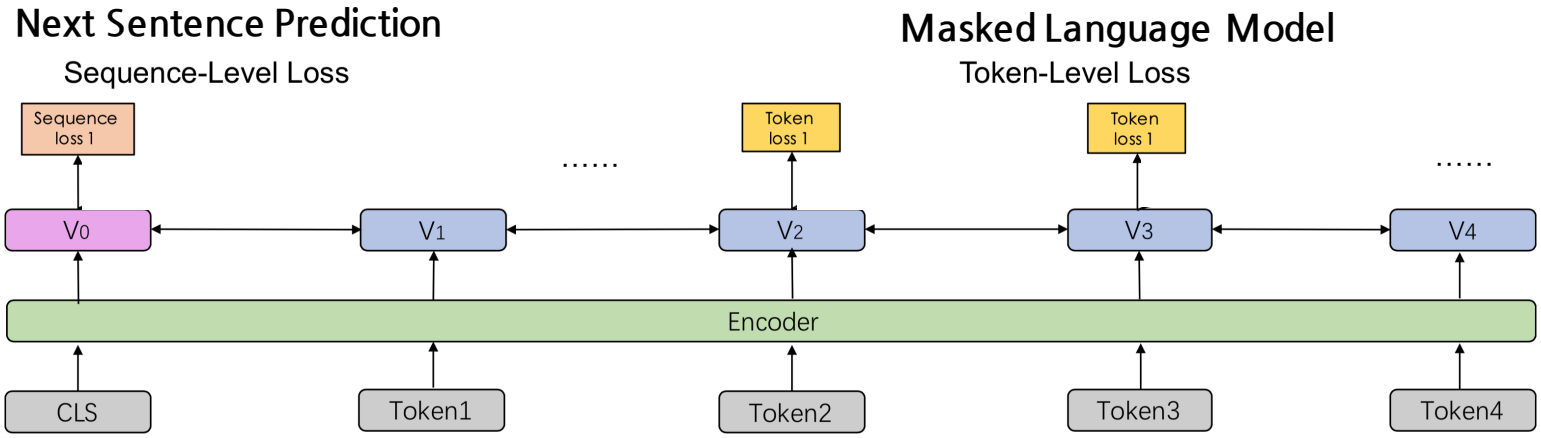
여러 task를 순차적으로 학습
단 새 task를 학습할 때 이전 task도 같이 학습해
이전 학습 내용을 잊어버리지 않도록 함



ERNIE2.0

- Continual Learning

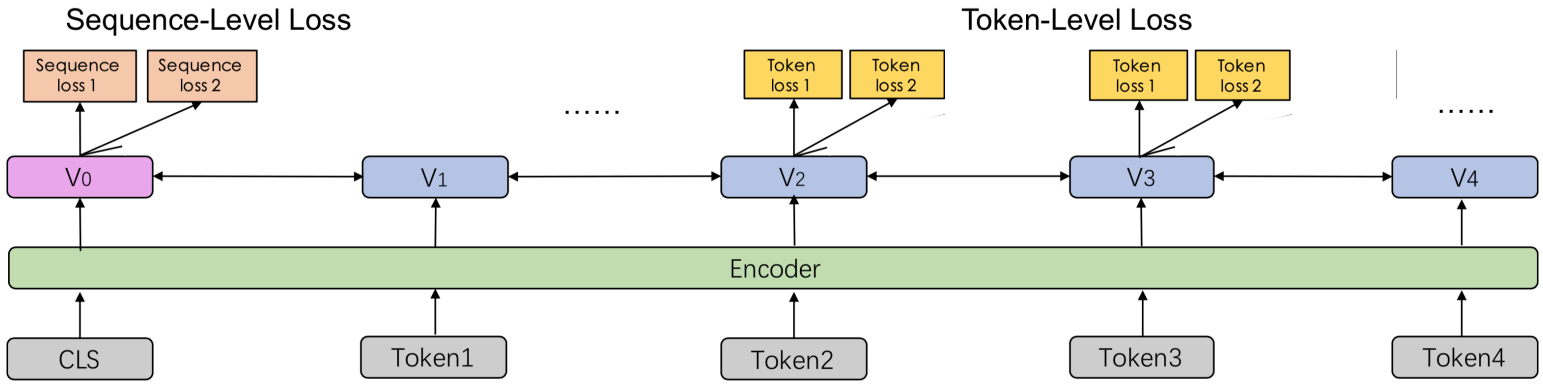
여러 task를 순차적으로 학습
단 새 task를 학습할 때 이전 task도 같이 학습해
이전 학습 내용을 잊어버리지 않도록 함



ERNIE2.0

- Continual Learning

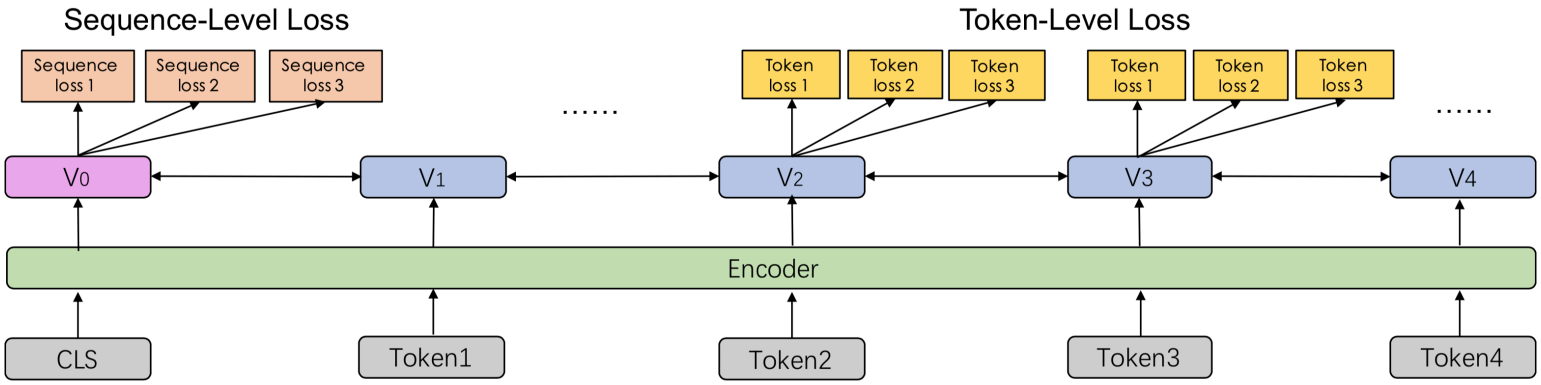
여러 task를 순차적으로 학습
단 새 task를 학습할 때 이전 task도 같이 학습해
이전 학습 내용을 잊어버리지 않도록 함



ERNIE2.0

- Continual Learning

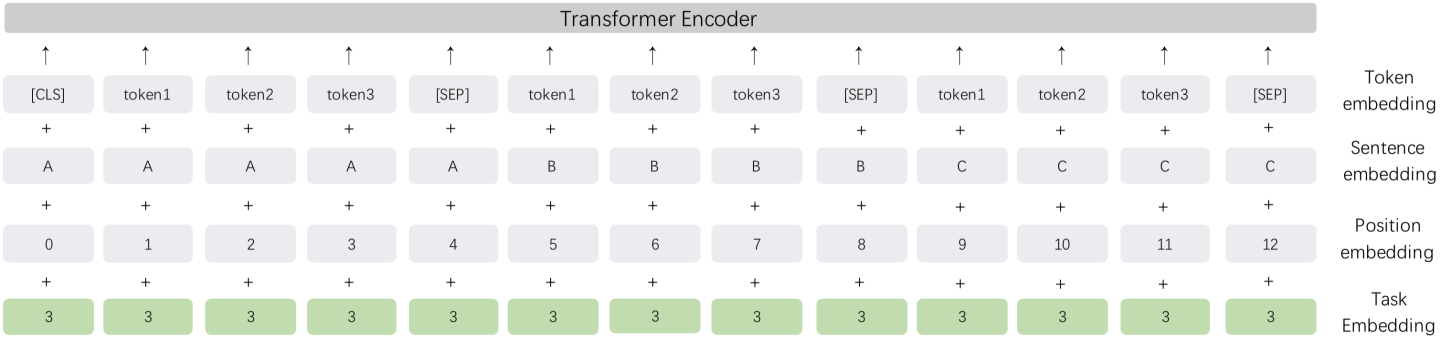
여러 task를 순차적으로 학습
단 새 task를 학습할 때 이전 task도 같이 학습해
이전 학습 내용을 잊어버리지 않도록 함



ERNIE2.0

- Task Embedding

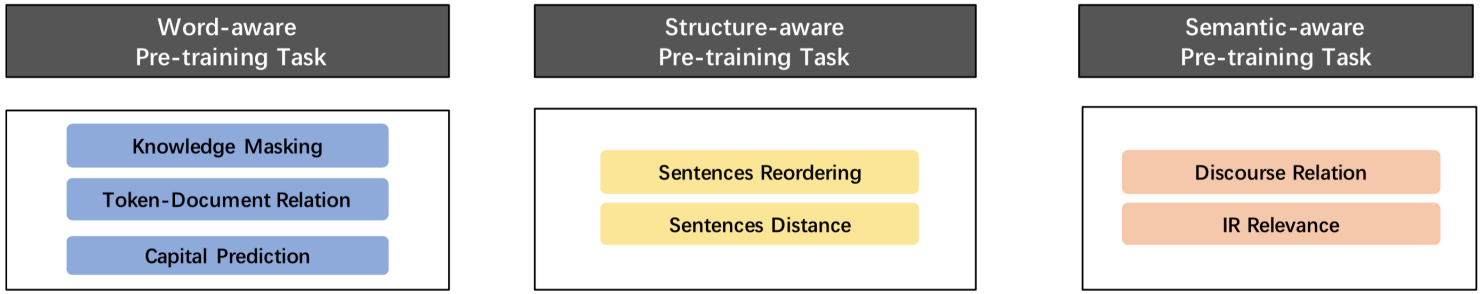
Task별 임베딩을 입력 레이어에서 더함



ERNIE2.0

- Pre-train Tasks

Word-aware Pre-training task：lexical information
Structure-aware Pre-training task：syntactic information
Semantic-aware Pre-training task：semantic information



ERNIE2.0

- Word-aware Pre-training task

Knowledge Masking Task
개체(named entity), 구(phrase) 단위 마스킹 (ERNIE1.0)
continual learning의 최초 모델에 적용

Capitalization Prediction Task
해당 단어가 대문자인지 소문자인지 예측

Token-Document Relation Prediction task
해당 토큰이 원본 문서의 다른 segment에 나타나는지 예측
키워드(동일 문서에 자주 나타나는 단어) 파악 가능

ERNIE2.0

- Structure-aware Pre-training task

Sentence Reordering Task
원래 세그먼트를 1~m개로 나누고 랜덤 셔플한 뒤 순서 맞추기
문장 간 관계를 파악하게 됨

ERNIE2.0

- Structure-aware Pre-training task

Sentence Reordering Task
원래 세그먼트를 1~m개로 나누고 랜덤 셔플한 뒤 순서 맞추기
문장 간 관계를 파악하게 됨

➔ m=2이라면 1! + 2! = 3, 즉 3범주 분류 문제로 정의
[1 0 0] : 세그먼트를 하나로 나눈 것
[0 1 0] : 두 세그먼트 가운데 현재 예측 대상 세그먼트가 첫번째

Sentence Distance task
3범주 분류 문제
0=동일 문서 앞뒤 문장
1=앞뒤는 아니지만 동일 문서
2=다른 문서에서 뽑힌 두 문장

ERNIE2.0

- Semantic-aware Pre-training task

Discource Relation Task
두 문장이 의미적으로, 비유적으로 유사한지 여부 예측

Sentence Distance task
앞 문장을 쿼리/뒤 문장을 검색 문서 제목으로 간주, 3범주 분류 문제
0=강한 관련성, 쿼리를 실제 검색했을 때 해당 문서 제목을 클릭
1=약한 관련성, 해당 쿼리를 던지면 검색은 되지만 미클릭
2=관련성 제로

ERNIE2.0

- Pre-train Data

English
Wikipedia, BookCorpus, Reddit, Discovery

Chinese
Encyclopedia, news, dialogue, 검색 등 (바이두 데이터)

ERNIE2.0

- Pre-train Settings

Layer : 12(base), 24(large)
Head : 12(base), 16(large)
Dimension : 768(base), 1024(large)
Adam optimizer : $\beta_1 = 0.9, \beta_2 = 0.98$
Batch size : 393216 tokens
Learning rate : 5e-5(English), 1.28e-4(Chinese)
Warm-up steps : 4000 (task 달라질 때마다)

ERNIE2.0

- Results (English)

Task(Metrics)	*BASE model*		*LARGE model*				
	Test		Dev			Test	
	BERT	ERNIE 2.0	BERT	XLNet	ERNIE 2.0	BERT	ERNIE 2.0
CoLA (Matthew Corr.)	52.1	**55.2**	60.6	63.6	**65.4**	60.5	**63.5**
SST-2 (Accuracy)	93.5	**95.0**	93.2	95.6	**96.0**	94.9	**95.6**
MRPC (Accurary/F1)	84.8/88.9	**86.1/89.9**	88.0/-	89.2/-	**89.7/-**	85.4/89.3	**87.4/90.2**
STS-B (Pearson Corr./Spearman Corr.)	87.1/85.8	**87.6/86.5**	90.0/-	91.8/-	**92.3/-**	87.6/86.5	**91.2/90.6**
QQP (Accuracy/F1)	89.2/71.2	**89.8/73.2**	91.3/-	91.8/-	**92.5/-**	89.3/72.1	**90.1/73.8**
MNLI-m/mm (Accuracy)	84.6/83.4	**86.1/85.5**	86.6/-	**89.8/-**	89.1/-	86.7/85.9	**88.7/88.8**
QNLI (Accuracy)	90.5	**92.9**	92.3	93.9	**94.3**	92.7	**94.6**
RTE (Accuracy)	66.4	**74.8**	70.4	83.8	**85.2**	70.1	**80.2**
WNLI (Accuracy)	**65.1**	**65.1**	-	-	-	65.1	**67.8**
AX(Matthew Corr.)	34.2	**37.4**	-	-	-	39.6	**48.0**
Score	78.3	**80.6**	-	-	-	80.5	**83.6**

ERNIE2.0

- Results (Chinese)

Task	Metrics	BERT$_{BASE}$		ERNIE 1.0$_{BASE}$		ERNIE 2.0$_{BASE}$		ERNIE 2.0$_{LARGE}$	
		Dev	Test	Dev	Test	Dev	Test	Dev	Test
CMRC 2018	EM/F1	66.3/85.9	-	65.1/85.1	-	69.1/88.6	-	**71.5/89.9**	-
DRCD	EM/F1	85.7/91.6	84.9/90.9	84.6/90.9	84.0/90.5	88.5/93.8	88.0/93.4	**89.7/94.7**	**89.0/94.2**
DuReader	EM/F1	59.5/73.1	-	57.9/72.1	-	61.3/74.9	-	**64.2/77.3**	-
MSRA-NER	F1	94.0	92.6	95.0	93.8	95.2	93.8	**96.3**	**95.0**
XNLI	Accuracy	78.1	77.2	79.9	78.4	81.2	79.7	**82.6**	**81.0**
ChnSentiCorp	Accuracy	94.6	94.3	95.2	95.4	95.7	95.5	**96.1**	**95.8**
LCQMC	Accuracy	88.8	87.0	89.7	87.4	**90.9**	**87.9**	**90.9**	**87.9**
BQ Corpus	Accuracy	85.9	84.8	86.1	84.8	86.4	85.0	**86.5**	**85.2**
NLPCC-DBQA	MRR/F1	94.7/80.7	94.6/80.8	95.0/82.3	95.1/82.7	95.7/84.7	95.7/85.3	**95.9/85.3**	**95.8/85.8**