

데이터와 알고리즘에 대한 이해를 기반으로 한 인공지능 활용 과정

강의부제 : 자연어처리 이론, 자연어처리 실습 예제

강의일 : 2023년 5월 25일

강연자: 주정헌 팀장 (와이엠에스닷코)



강의 목차

5월 9일 (1일차) : 인공지능의 발달 및 역사, 인공지능의 연구분야

5월 11일 (2일차) : 인공지능 알고리즘 종류, 정형/비정형 데이터, 산업별 활용

5월 16일 (3일차) : 파이썬 설치, 문법, 자주 사용되는 라이브러리 활용 방법

5월 18일 (4일차) : 딥러닝 기초 설명, 파이썬 및 Colab 실습

5월 23일 (5일차) : 파이썬 및 Colab 실습, 이미지 인식 실습 예제

5월 25일 (6일차) : 자연어처리 이론, 자연어처리 실습 예제

당일 목차

1. 자연어 처리 기초 설명
2. 자연어 이해 기초 설명
3. 자연어 생성 기초 설명
4. 자연어 처리 실습 예제

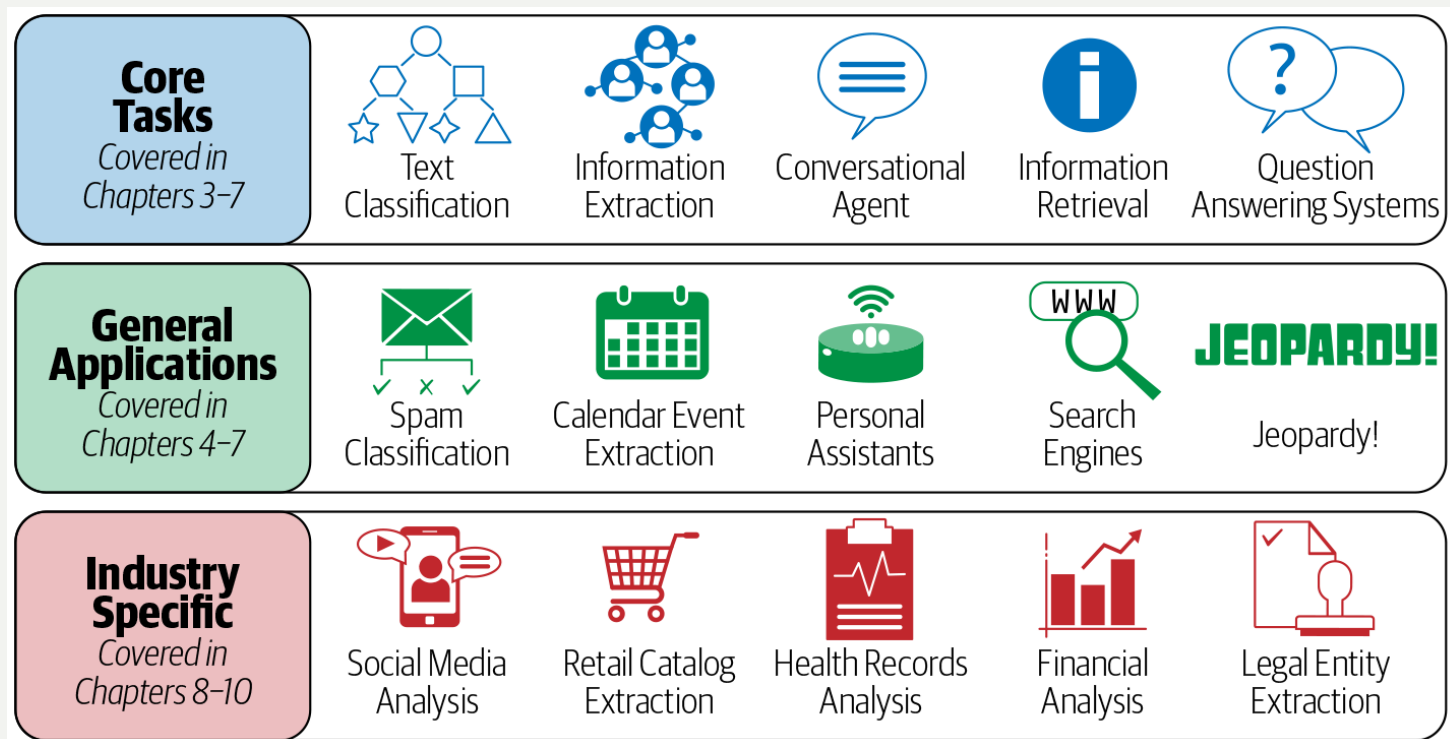
1. 자연어처리 기초

자연어처리란?

NLP의 주요 목표는 컴퓨터가 의미있고 유용한 방식으로 **인간 언어와 상호작용**할 수 있도록 하는 것. 이를 위해 알고리즘, 모델, 기술을 개발하여 자연어 텍스트나 음성을 분석, 이해 및 생성을 하는 기술을 개발

1. 자연어처리 기초

자연어처리 활용 영역



2. 자연어처리 이해 기초

자연어이해(NLU, Natural language understanding)

NLU는 기계가 인간 언어의 의미, 문맥, 의도를 이해할 수 있도록 하는 것을 목표로 함

1. 구문분석(Syntax Analysis) : 문장의 문법 구조를 구문 분석하여 단어와 구문이 어떻게 관련되는지 이해하는 과정
2. 의미분석(Semantic Analysis) : 단어와 구문 간의 관계를 이해하여 텍스트에서 의미를 추출하는 과정에 초점
3. 화용분석(Pragmatic Analysis) : 언어 이해의 문맥적 및 상황적 측면을 고려

2. 자연어처리 이해 기초

BERT 모델 - 입력 데이터

Input	[CLS]	my	dog	is	cute	[SEP]	he	likes	play	##ing	[SEP]
Token	$E_{[CLS]}$	E_{my}	E_{dog}	E_{is}	E_{cute}	$E_{[SEP]}$	E_{he}	E_{likes}	E_{play}	$E_{##ing}$	$E_{[SEP]}$
	+	+	+	+	+	+	+	+	+	+	+
Segment	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
	+	+	+	+	+	+	+	+	+	+	+
Position	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}

1. Token Embedding

가장 자주 등장하는 단위를 최소의 형태로 쪼갬
것을 토큰이라 부름

2. Segment Embedding

문장들간의 문장들을 1과 0으로 반복하여
구분하는 것

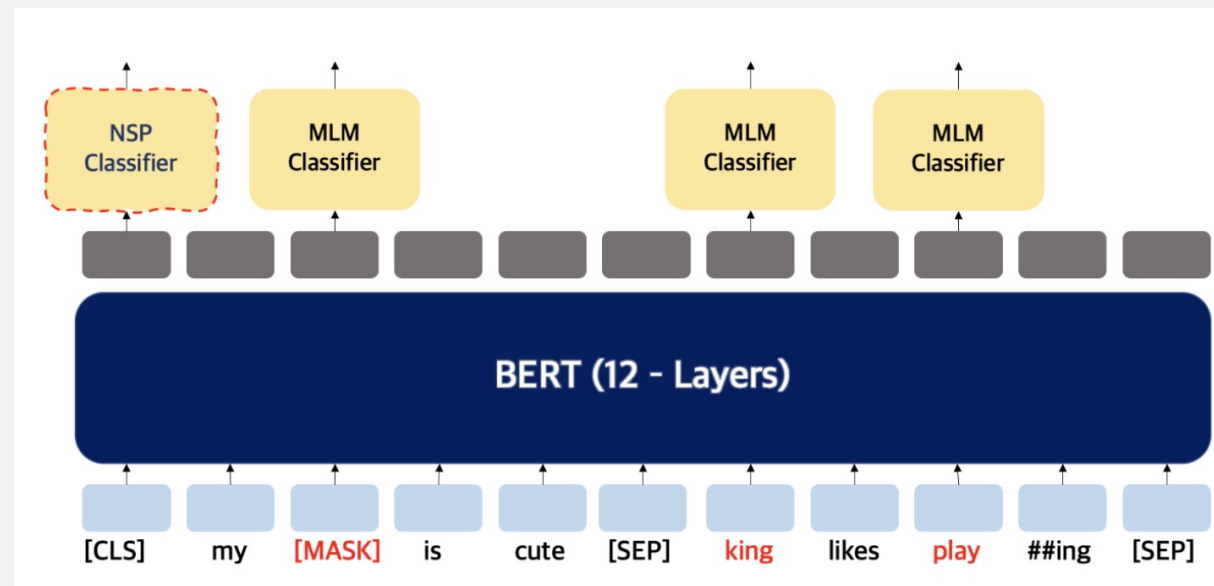
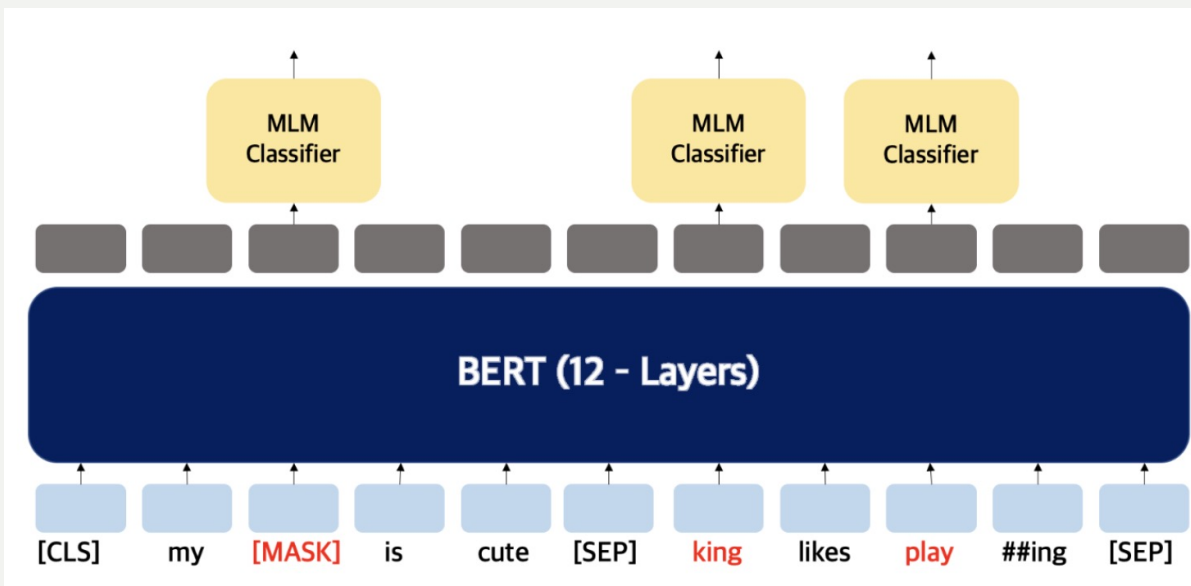
3. Position Embedding

Self-attention으로 계산할 때 각 단어는 중요도를
모르기 때문에 각 단어들의 순서를 기입하여 위치
정보를 알려줌

2. 자연어처리 이해 기초

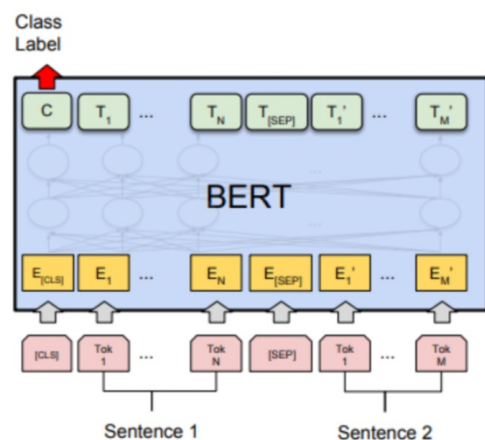
BERT 모델 - 학습방법 MLM & NSP

Masked LM & Next Sentence Prediction

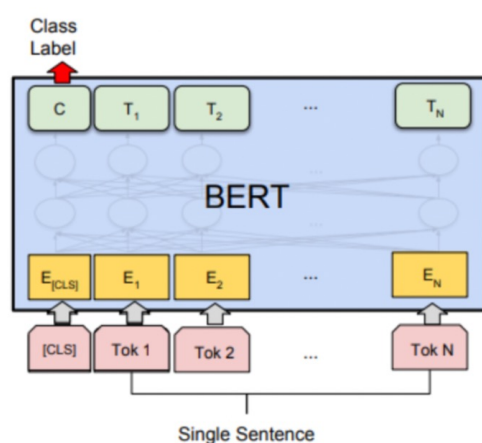


2. 자연어처리 이해 기초

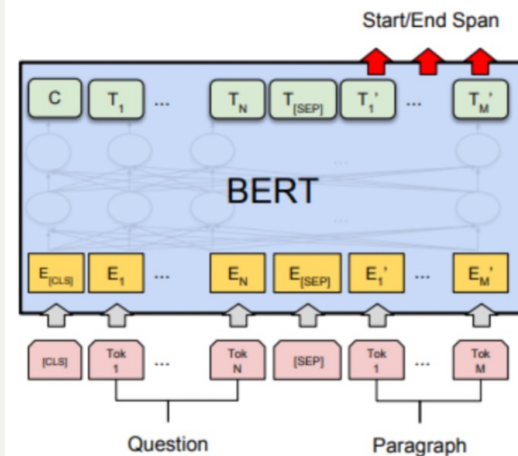
BERT 모델 - 미세조정 방법



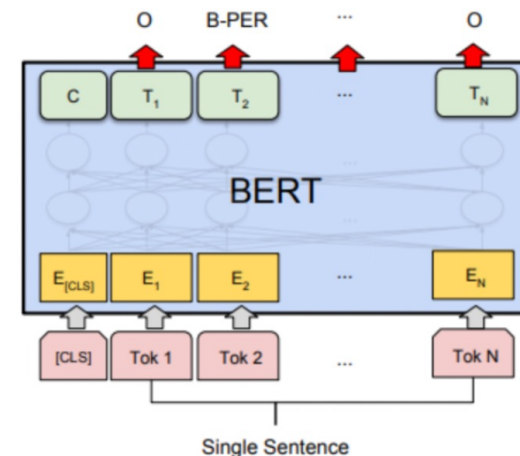
(a) Sentence Pair Classification Tasks:
MNLI, QQP, QNLI, STS-B, MRPC,
RTE, SWAG



(b) Single Sentence Classification Tasks:
SST-2, CoLA



(c) Question Answering Tasks:
SQuAD v1.1



(d) Single Sentence Tagging Tasks:
CoNLL-2003 NER

3. 자연어처리 생성 기초

자연어생성(NLU, Natural language generation)

기계에 의해 사람과 유사한 언어를 자동으로 생성하는 분야. 일관되고 의미 있는 문장, 단락 또는 전체 문서를 생성할 수 있는 시스템을 개발하는 것을 포함

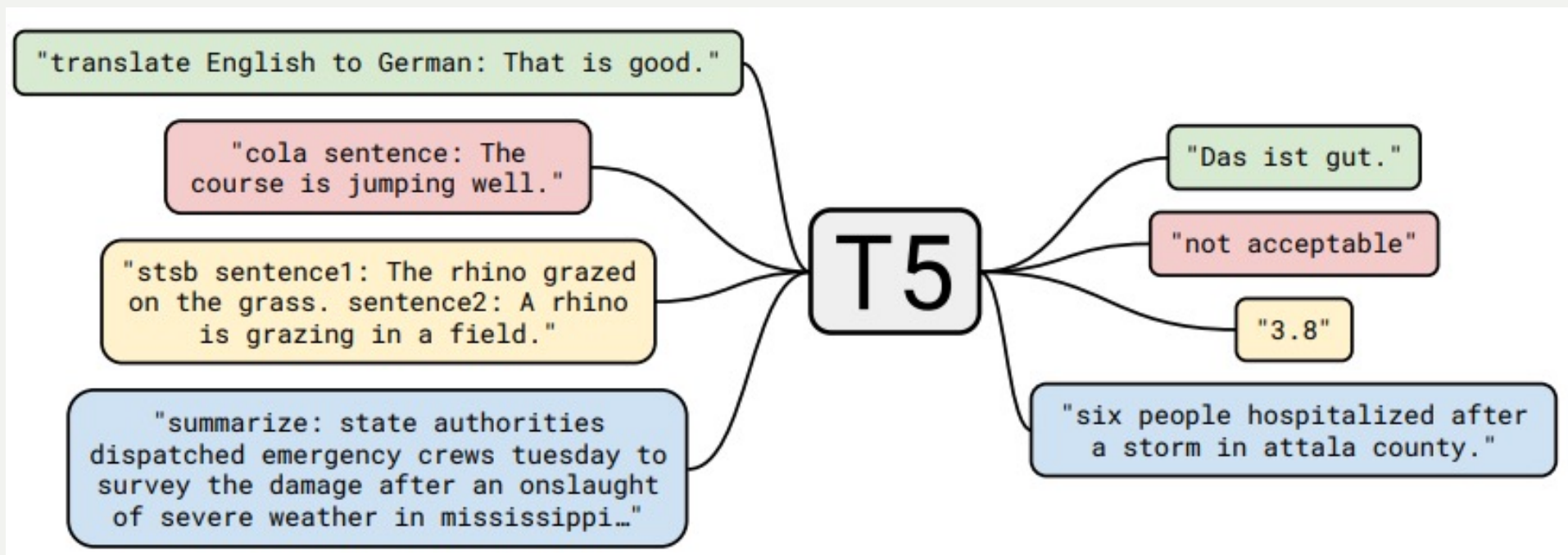
문서 계획(Text Planning): 텍스트 계획은 생성된 텍스트의 전반적인 구조와 내용을 결정하는 작업. 주요 내용을 선택하고 정보를 조직화하며 적절한 세부 수준을 결정

콘텐츠 선택(Content Selection): 콘텐츠 선택은 생성된 텍스트에 포함할 관련 정보를 선택하는 작업

문장 생성(Sentence Generation): 문장 생성은 문법적으로 정확하고 일관된 문장을 생성하는 작업에 중점

3. 자연어처리 생성 기초

자연어생성 모델 - T5



3. 자연어처리 생성 기초

자연어생성 모델 - T5의 학습 방법

Denoising Corrupted Span

Original text

Thank you ~~for inviting~~ me to your party ~~last~~ week.

Inputs

Thank you <X> me to your party <Y> week.

Targets

<X> for inviting <Y> last <Z>

4. 자연어처리 실습 예제

Colab을 열어보세요~



감사합니다!

