

교육 과정 소개서.

Pytorch를 활용한 자연어처리 심화 CAMP



코스요약

코스명	Pytorch를 활용한 자연어처리 심화 CAMP
기간	2019. 11. 16 ~ 2020. 1. 18 (12/14, 12/21, 12/28 휴강)
일정	매주 토요일 10:00 -13:00
장소	패스트캠퍼스 강남강의장
준비물	개인 노트북
담당자	02-501-9396 / help-ds@fastcampus.co.kr
수강료	1,500,000
상세페이지 url	https://www.fastcampus.co.kr/data_camp_nlbady/

코스목표

기본적인 Classification을 넘어 번역, 요약, 문장 생성에 필요한알고리즘을 이해하고, 언어 모델등을 구현하는 실습을 통해상용화 가능한 수준의 자연어처리 역량을 기릅니다.

코스정보

본 강의는 NLP의 정점에 있는 최신 모델에 대한 깊이있는 커리큘럼 구성과 실무를 방불케하는 실습으로 구성되어 있습니다. 이 모든 내용은 기계번역, NLP의 분야의 전문가이신 현업 강사님이 진행하시기 때문에 직접 해보지 않으면 알 수 없는 생생한 노하우도 함께 얻어가실 수 있습니다.



코스특징

SEQ2SEQ 모델 완벽 이해

자연어처리 분야의 핵심 딥러닝 기술인 Seque2Seq 모델과 관련된 이론에 대해 깊이있게 다룹니다. 또한 기계번역, NLP의 분야에서 전문가이신 강사님과 함께 실제 상품화가 가능한 수준의 실습을 통해 seq2seq의 실무 활용 방법에 대해 익힙니다.

NLG의 근간, 언어모델 구현

language Modeling은 그 자체로서 활용도 뿐만 아니라, 기계번역과 음성인식, OCR, 그리고 Natural Language Generation에 있어서도 매우 중요한 역할을 차지 합니다. 본 강의에서는 언어모델의 이론에 대해 깊이있게 이해하고, 직접 언어 모델링을 실습합니다.

자연어 처리에 강화학습 적용

본 강의에서는 Non-differentiable objective function 이나 Adversarial learning 활용한 NLP 성능 향상과 관련된 강화학습에 대해 이론과 실습을 통해 가르쳐드립니다.

필독! 수강 전 지식.

아래 내용에 2개 이상 해당하는 분들이 원활하게 들으실 수 있는 수업입니다.
자세한 상담을 원하시면 담당 매니저에게 연락해주세요.

- ㉔ PyTorch, TensorFlow, Keras 등의 딥러닝 프레임워크로 간단한 프로젝트를 해본 적이 있다. (e.g. 텍스트 분류)
- ㉕ 기본적인 자연어처리(Word Embedding, Text Classification 등)에 대한 내용을 알고 있다.
- ㉖ LSTM과 GRU 사이의 차이점과 기존 RNN과의 차이점을 말할 수 있다.
- ㉗ 연속 랜덤 변수와 이산 랜덤 변수의 차이를 말할 수 있다.
- ㉘ 조건부 확률 분포와 주변 확률 분포(marginal probability distribution)에 대해 알고 있다.



커리큘럼

- 1주차 ● 자연어처리 소개 & 자연어처리 전처리**
 - 세션 1: 자연어처리의 흐름을 살펴보고 파이토치를 활용한 간단한 실습
 - 세션 2: 자연어처리 실무에서 가장 중요한 전처리 (수집과 정제, 문장 분절과 단어 분절, BPE 분절, 병렬 코퍼스 만들기)

- 2주차 ● 언어모델에 대한 이해 & 기계번역을 통한 Seq2Seq 이해**
 - 세션 3: 언어 모델 소개, n-gram과 perplexity, 희소성(sparsity), 뉴럴네트워크 언어 모델링, 활용 분야 소개
 - 세션 4: 기계번역 소개, Sequence-to-Sequence 이해하기, Attention 깊게 제대로 이해하기, Auto-regressive 속성과 Teacher-forcing 이해하기

- 3주차 ● 트랜스포머를 통한 성능 고도화 & 기계번역 심화 주제 학습**
 - 세션 5, 6: Multi-head Attention, Self-Attention과 Masking, Block을 쌓아 Encoder & Decoder 만들기, Position Encoding

- 4주차 ● 기계번역 실습**
 - 세션 7: Sequence to Sequence Exercise 코드 실습, Transformer 코드 실습, Mini-batch Parallelized Beam-search 코드 실습
 - (세션 7의 실습은 5주차에도 일부 이어서 진행합니다)

- 5주차 ● 강화학습을 활용한 자연어 생성**
 - 세션 8: 강화학습 소개, 폴리시 그라디언트 (Policy Gradient) 소개, 자연어생성에서의 강화학습의 특성, 강화학습을 활용한 자연어 생성 성능 극대화 방법 소개

- 6주차 ● 강화학습을 활용한 자연어 생성 실습 & 듀얼리티(Duality)를 활용한 자연어 생성**
 - 세션 9: Minimum Risk Training (MRT) 실습
 - 세션 10: Duality 개념 소개, Dual Supervised Learning (DSL) 소개, Dual Unsupervised Learning (DUL) 소개, Back Translation의 재해석

- 7주차 ● 듀얼리티(Duality)를 활용한 자연어 생성 실습 & 요약 및 정리**
 - 세션 11: Dual Supervised Learning 실습
 - 세션 12: 7주 동안의 강의에 대한 전체 요약 & 정리



강사소개



김기현

[약력]

- 現 마키나락스 수석연구원 (이상탐지 및 생성 모델 학습 연구/개발)
- 前 SK플래닛 머신인텔리전스랩 (신경망 기계번역 연구/개발 및 상용화, SK플래닛 SUPEX 수상)
- 前 티켓몬스터 데이터랩 (추천시스템 및 자연어처리 시스템 개발 및 상용화)
- 前 한국전자통신연구원(ETRI) 자동 통역 연구실 (자동 통역(음성인식 및 언어모델) 연구/개발 및 상용화)
(특허기술상(특허청) 충무공상 : 자동통역시스템 발명)
- Qualcomm R&D Korea 인턴 및 표창
- Stony Brook University 컴퓨터공학과 학사/석사
- 한국정보올림피아드(KOI) 19회 동상, 20회 장려상
- 머신러닝 관련 다수 특허 등록

[강의 경험]

- 팟캐스트 캠퍼스 PyTorch를 활용한 자연어처리 심화 CAMP 강의
- 팟캐스트 캠퍼스 자연어처리를 위한 딥러닝 CAMP
- KT 및 KTDS 자연어처리 강의
- SK그룹 기계번역 강의
- 유튜브 모두의 딥러닝 시즌2 PyTorch 강의 참여



수강환경

패스트캠퍼스 강남강의장



❖ 강의에 따라 강의장이 변경될 수 있습니다.