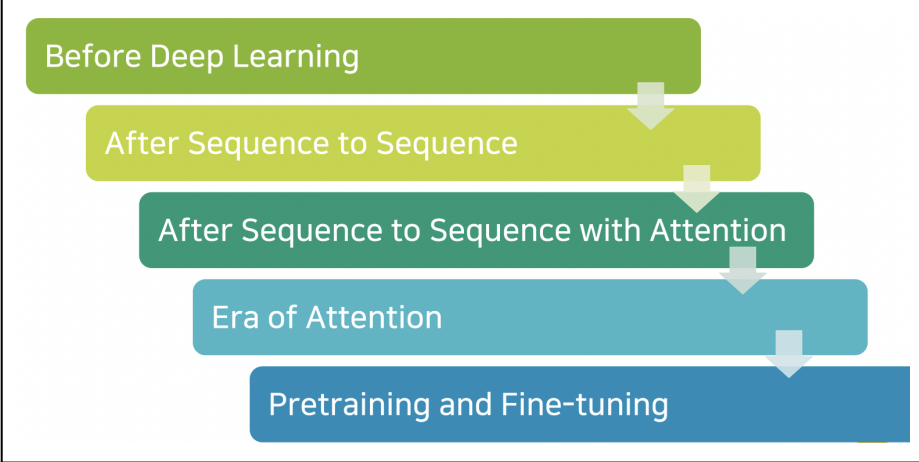


교육 제목	데이터 기반 인공지능 시스템 엔지니어 양성 과정
교육 일시	2021년 11월25일
교육 장소	YGL C-6 학과장 & 자택(디스코드 이용한 온라인)
교육 내용	
오전	<p>자연어 처리</p> <p>1. 자연어 처리란?</p> <ul style="list-style-type: none"> • 자연어(Natural Lanaguage)란? <ul style="list-style-type: none"> • 자연어 혹은 자연 언어는 사람들이 일상적으로 쓰는 언어를 인공적으로 만들어진 언어인 인공어와 구분하여 부르는 개념 (출처: 위키피디아) • Natural Language Processing <ul style="list-style-type: none"> • 사람이 이해하는 자연어를 컴퓨터가 이해할 수 있는 값으로 바꾸는 과정(NLU) • 더 나아가 컴퓨터가 이해할 수 있는 값을 사람이 이해하도록 바꾸는 과정(NLG) <p>2. NLP vs Others</p> <ul style="list-style-type: none"> • Computer Vision <ul style="list-style-type: none"> • Image Recognition • Object Detection • Image Generation • Super Resolution • Natural Language Processing <ul style="list-style-type: none"> • Text Classification • Machine Translation • Summarization • Question Answering • Speech Processing <ul style="list-style-type: none"> • Speech Recognition (STT) • Speech Synthesis (TTS) • Speaker Identification • Reinforcement Learning <p>3. NLP 와 Deep Learning</p> <p>이산적인 심볼릭 데이터 - 사람이 이해할 수 있음 - 입력: x - 출력: y</p> <p>$h_x = f_e(x; \theta_e)$, 심볼릭→연속데이터 by 임베딩 계층 / 인코더</p> <p>$\frac{\partial \mathcal{L}}{\partial \theta_e}$</p> <p>$\frac{\partial \mathcal{L}}{\partial \theta_r}$</p> <p>$\frac{\partial \mathcal{L}}{\partial \theta_d}$</p> <p>$\mathcal{L}(\theta) \propto \text{Error}(y, y^*)$</p> <p>$y = f_d(h_y; \theta_d)$, 연속데이터→심볼릭 by 생성 계층 / 디코더</p> <p>연속적인 데이터 - 연산 효율 높음 - 입력: h_x - 출력: h_y</p> <p>$h_y = f_r(h_x; \theta_r)$, 인코더와 디코더</p>

	<p>4. NLP의 어려움</p> <ul style="list-style-type: none"> • 이산 값을 갖는 자연어는 사람의 입장에서 인지가 쉬울 수 있으나, 기계의 입장에서 매우 어려운 값 • One-hot encoding으로 표현된 값은 유사도나 모호성을 표현할 수 없다. <ul style="list-style-type: none"> • 서로 다른 One-hot 벡터끼리의 유사도나 거리는 모두 동일하다. • 따라서, 아래의 질문에 대답할 수 없다. <ul style="list-style-type: none"> • <파랑>과 <핑크> 중에서 <빨강>에 가까운 단어는 무엇인가요? • 하지만, 사람의 어휘 체계는 계층적 구조를 띄고 있다. • 또한 높은 차원으로 표현되어 매우 sparse하게 된다. <p>딥러닝에서는 Word Embedding을 통해 해결</p> <p>5. 한국어의 어려움</p> <ul style="list-style-type: none"> • 접사를 사용하는 교착어 (+ 유연한 어순) • 띄어쓰기 • 평서문, 의문문의 구분 • 주어의 부재 • 한자 기반의 언어 • 단어 중의성 <p>6. NLP의 역사(?)</p>  <pre> graph TD A[Before Deep Learning] --> B[After Sequence to Sequence] B --> C[After Sequence to Sequence with Attention] C --> D[Era of Attention] D --> E[Pretraining and Fine-tuning] </pre>
오후	별도 실습 파일 참조