

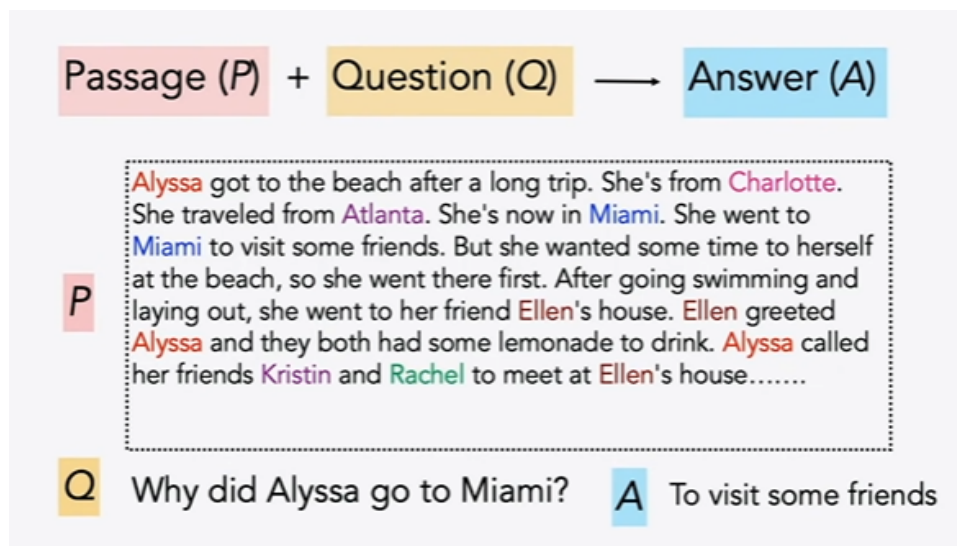
Week 11

☰ 태그	
☰ 열	

Question Answering

Motivation & History

- 질문에 대한 답을 얻기를 원할 때
 1. Finding documents that contain an answer
 - a. information retrieval
 2. Finding an answer in a paragraph or a document
 - a. Machine Learning Comprehension = MRC
- MCTest Reading Comprehension
 - simple MC challenge



Stanford Question Answering Dataset (SQuAD)

- 잘 구성된 dataset

- **SQuAD Limitations**
 - Span-based answers 만 존재 (no yes/no, counting, implicit why)
 - Passages 내에서만 정답을 찾으려 하는 질문 구성
 - 여러 문서들을 비교하여, 진짜 정답을 찾아낼 필요가 없음
 - 실제 마주하게 될 질문-답변 보다 잘 정립된 문법 구조
 - 동일 지시어(coreference) 문제를 제외하고는 Multi-fact 문제, 문장 추론 문제가 거의 없음
- Nevertheless, SQuAD is a well-targeted, well-structured, clean dataset
 - QA 문제를 푸는데 있어 가장 많이 사용되고, 경쟁하고 있는 데이터셋
 - 실제 시스템을 개발하기 위한 유용한 starting point

QA model

- Stanford Attentive Reader
 - 질의에 대한 응답을 찾는 모델을 구축하는데 BiLSTM with attention을 사용
 - Question에 대한 vector 생성 → Passage에 대한 vector 생성 → Attention
- Stanford Attentive Reader++
 - 3 layer BiLSTM
 - BiLSTM의 hidden state를 포지션별로 concat후, weighted sum을 구하여 q벡터 구성함.
 - Glove 벡터만 사용한 것이 아닌, 단어의 feature도 같이 넣어줌.
- BiDAF: Bi-Directional Attention Flow for Machine Comprehension
 - Attention이 양방향으로 적용됨 (question ↔ paragraph)
 1. Character Embedding Layer: CharCNN을 사용하여 각단어를 vector space에 mapping 함.
 2. Word Embedding Layer: pre-trained word embedding 모델을 사용하여 각 단어를 vector space에 mapping 함
 3. Contextual Embedding Layer: Target word의 주변단어들을 통해 embedding 을 정제함. 처음 3개의 layer에 대해서는 query 와 context에 모두 적용됨.
 4. Attention Flow Layer: Query 와 Context 를 쌍으로 묶어서 Attention 을 학습하게 된다.
 5. Modeling Layer: RNN을 통해 Context를 탐색함.
 6. Output Layer: Query에 대해 답을 생성함.
 - 양방향 attention을 가능하게 하기 위해 shared matrix인 S 사용함.

