

## BI-DIRECTIONAL ATTENTION FLOW FOR MACHINE COMPREHENSION

week 12 복습과제 권재선

- MC(Machine Comprehension) : answering a query about a given context paragraph
  - 문장(context)의 흐름을 이해하고, 주어진 문장(context)에서 질문(query)에 대답하는 것
  - query와 context의 복잡한 상호관계를 이해해야함
- 최근에 Attention 모델이 MC에 사용되기 시작함
  - 문맥의 작은 부분에 집중 가능하며, 그 정보를 fixed-size vector로 표현
- BiDAF : Bi-Directional Attention Flow
  - multi-stage hierarchical process
  - context를 요약하지않고 query-aware context vector를 찾아내는 방식
  - context를 세분화시킴(query에서 사용되는 중요한 단어를 골라냄)
- query-aware context vector로 context summarization 대체
  - context가 query와 연관지어져 query에 필요한 특성을 반영한 vector

### BIDAF 장점

1. attention layer가 문맥을 fixed-size vector로 요약하는 것으로 사용되지 않음.
  - 이전 layer에서 계산된 attended vector를 subsequent model로 flow하게 해줌
  - 이 방법을 통해 요약으로 인한 정보손실이 줄어들게 됨
2. memory-less attention
  - time-step마다 attention을 계산하여 이전 attention layers를 사용하지않음
  - 따라서 현재 time step에서 query와 context의 attention을 학습하는 것에 집중
  - 또한, 이전의 잘못된 attention의 영향을 받지 않게 됨
3. bi-directional
  - context와 query가 충분히 정보를 교환가능

### Architecture

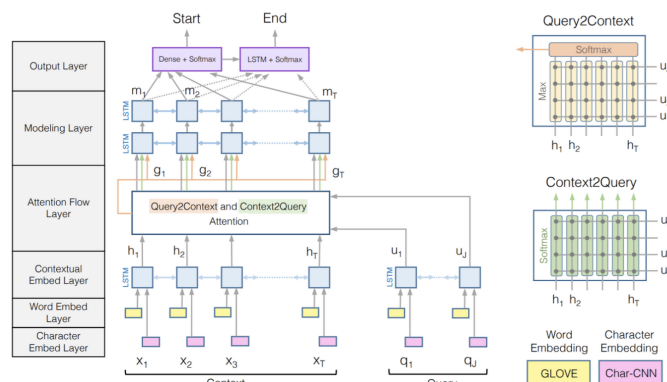


Figure 1: BiDirectional Attention Flow Model (best viewed in color)

## <BiDAF 기본구조>

- Character Embedding Layer: CharCNN을 사용하여 각 단어를 vector space에 mapping
- Word Embedding Layer: pre-trained word embedding 모델을 사용하여 각 단어를 vector space에 mapping
- Contextual Embedding Layer: Target word의 주변 단어들을 통해 embedding을 정제, 처음 3개의 Layer에 대해서는 Query와 Context에 모두 적용
- Attention Flow Layer: Context에 대해 Query-aware feature vector를 만들기 위해 Query와 Context를 쌍으로 묶어 Attention을 학습하게 됨
- Modeling Layer: RNN을 통해 Context를 탐색
- Output Layer: Query에 대해 답을 생성

### <Character Embedding Layer>

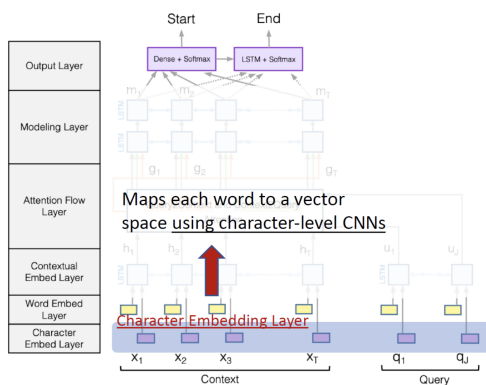


Figure 1: BiDirectional Attention Flow Model (best viewed in

### <Word Embedding Layer>

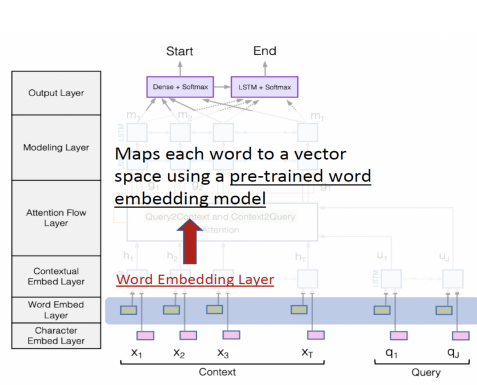


Figure 1: BiDirectional Attention Flow Model (best viewed in

### <Contextual Embedding Layer>

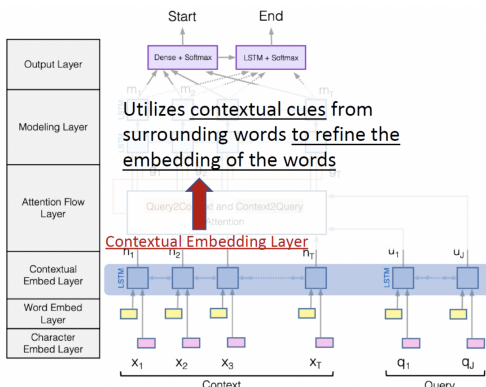


Figure 1: BiDirectional Attention Flow Model (best viewed in

### <Attention flow layer>

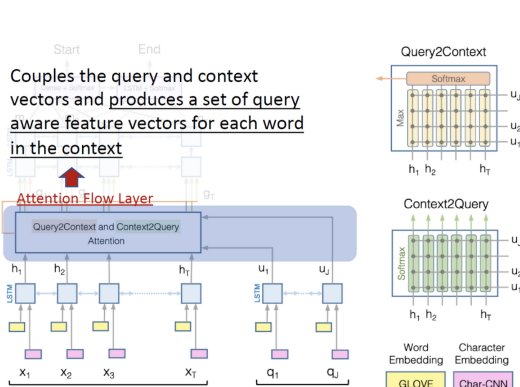
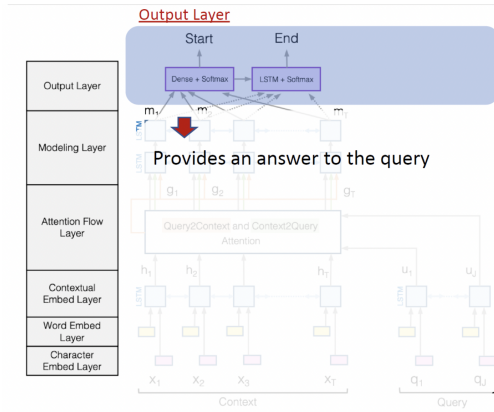
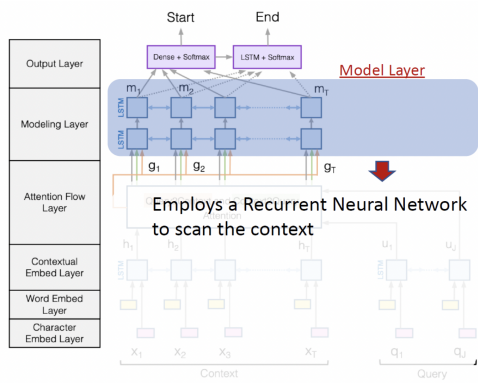


Figure 1: BiDirectional Attention Flow Model (best viewed in color)

### <Modeling Layer>

### <Output Layer>



## <Result>

	Single Model		Ensemble	
	EM	F1	EM	F1
Logistic Regression Baseline <sup>a</sup>	40.4	51.0	-	-
Dynamic Chunk Reader <sup>b</sup>	62.5	71.0	-	-
Fine-Grained Gating <sup>c</sup>	62.5	73.3	-	-
Match-LSTM <sup>d</sup>	64.7	73.7	67.9	77.0
Multi-Perspective Matching <sup>e</sup>	65.5	75.1	68.2	77.2
Dynamic Coattention Networks <sup>f</sup>	66.2	75.9	71.6	80.4
R-Net <sup>g</sup>	<b>68.4</b>	<b>77.5</b>	72.1	79.7
BiDAF (Ours)	68.0	77.3	<b>73.3</b>	<b>81.1</b>

(a) Results on the SQuAD test set

	EM	F1
No char embedding	65.0	75.4
No word embedding	55.5	66.8
No C2Q attention	57.2	67.7
No Q2C attention	63.6	73.7
Dynamic attention	63.5	73.6
BiDAF (single)	67.7	77.3
BiDAF (ensemble)	72.6	80.7

(b) Ablations on the SQuAD dev set