Bi-Directional attention flow for machine comprehension

1. introduction

In this paper, we introduce the Bi-Directional Attention Flow (BIDAF) network, a hierarchical multi-stage architecture for modeling the representations of the context paragraph at different levels of granularity. While we iteratively compute attention through time as in Bahdanau et al. (2015), the attention at each time step is a function of only the query and the context paragraph at the current time step and does not directly depend on the attention at the previous time step. It forces the attention layer to focus on learning the attention between the query and the context, and enables the modeling layer to focus on learning the interaction within the query-aware context representation (the output of the attention layer).

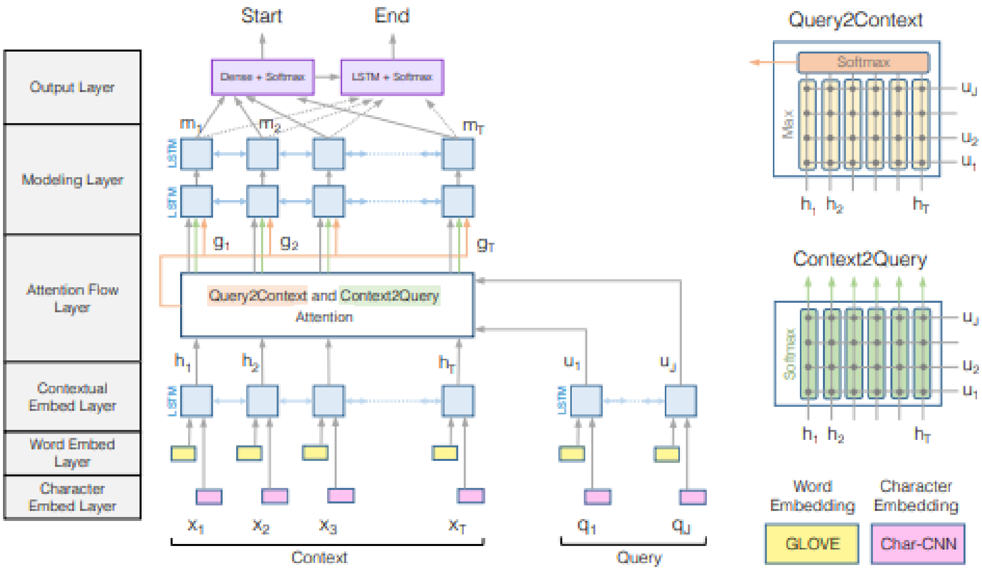

Figure 1: BiDirectional Attention Flow Model *(best viewed in color)*

2. Model

모델은 6개 layer로 구성된다.

- Character Embedding Layer maps each word to a vector space using character-level CNNs.

각단어를 고차원 벡터 공간에 매핑하는 역할

- Word Embedding Layer maps each word to a vector space using a pre-trained word embedding model.

사전 훈련된 단어 벡터인 GloVe를 사용하여 각 단어의 고정 단어 임베딩을 얻는다.

- Contextual Embedding Layer utilizes contextual cues from surrounding words to refine the embedding of the words. These first three layers are applied to both the query and context.

- Attention Flow Layer couples the query and context vectors and produces a set of queryaware feature vectors for each word in the context.

Context 및 query 단어 정보를 연결하고 융합하는 역할.

$$\mathbf{S}_{tj} = \alpha(\mathbf{H}_{:t}, \mathbf{U}_{:j}) \in \mathbb{R}$$

- Modeling Layer employs a Recurrent Neural Network to scan the context.

- Output Layer provides an answer to the query. $\mathbf{p}^1 = \mathrm{softmax}(\mathbf{w}^\top_{(\mathbf{p}^1)}[\mathbf{G};\mathbf{M}])$, $\mathbf{p}^2 = \mathrm{softmax}(\mathbf{w}^\top_{(\mathbf{p}^2)}[\mathbf{G};\mathbf{M}^2])$