Week12_복습과제

BI-DIRECTIONAL ATTENTION FLOW FOR MACHINE COMPREHENSION
논문 요약과제

- Machine comprehension(MC): answering a query about a given context paragraph , requires modeling complex interactions between the context and the query.

1. Inroduction
: MC - Systems trained end-to-end which enables the system to focus on a targeted area within a context paragraph
: Attention mechanisms - the computed attention weights are often used to extract the most relevant information from the context for answering the question by summarizing the context into a fixed-size vector.
: in the text domain, they are often temporally dynamic, whereby the attention weights at the current time step are a function of the attended vector at the previous time step.
: they are usually uni-directional, wherein the query attends on the context paragraph or the image.

: Bi-Directional Attention Flow network, BIDAF: a hierarchical multi-stage architecture for modeling the representations of the context paragraph at different levels of granularity. it includes character-level, word-level, and contextual embeddings, and uses bi-directional attention folw to obtain a query-aware context representation.
: Our attention layer is not used to summarize the context paragraph into a fixed-size vector.
: is computed for every time step and the attended vector at each time step.
: is allowed flow - reduces the information loss caused by early summarization.
: We use a memory-less attention mechanism.
: does not directly depend on the attention at the previous time step.
: leads to the division of labor between the attention layer and the modeling layer.
: query-aware context representation (the output of the attention layer).
: we use attention mechanism in both directions, query-to-context and context-to-query, which provide complimentary information to each other.

Week12_복습과제

2. Model

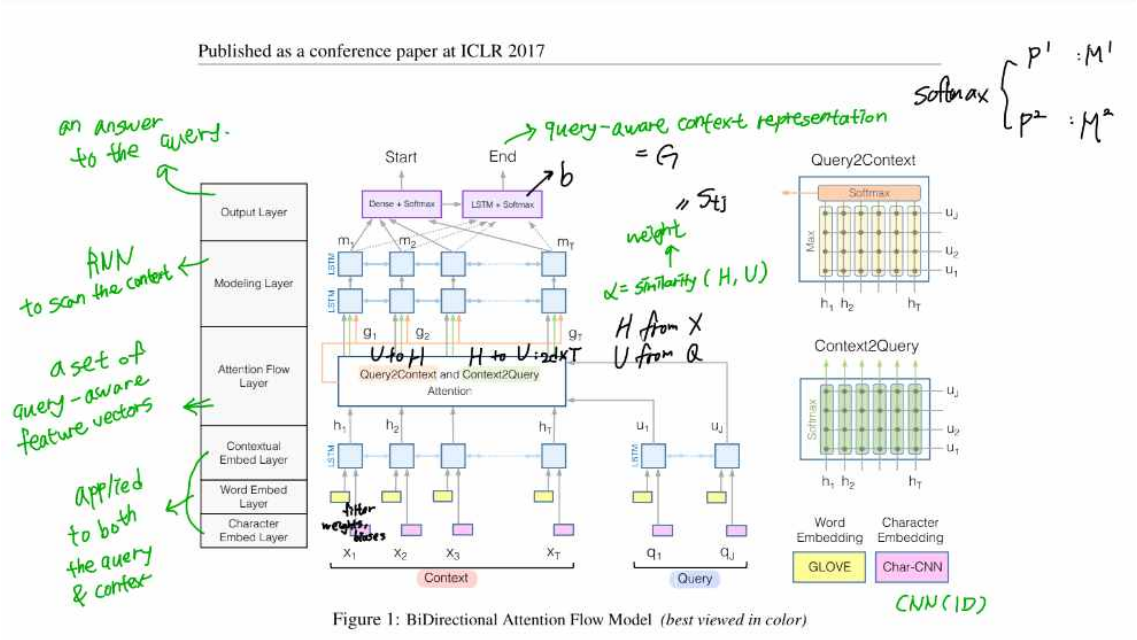

Figure 1: BiDirectional Attention Flow Model *(best viewed in color)*

-1. Character Embedding Layer
: maps each word to a vector space using character-level CNNs

-2. Word Embedding Layer
: maps each word to a vector space using a pre-trained word embeddin model.(GLoVe)

-3. Contextual Embedding Layer
: utilizes contextual cues from surrounding words to refine the embedding of the words. These first layers are applied to both the query and context.

-4. Attention Flow Layer
: couples the query and context vectors and produces a set of query-aware feature vectors for each word in the context.

-5. Modeling Layer
: employs a Recurrent Neural Network to scan the context.

-6. Output Layer
: provides an answer to the query.

Week12_복습과제

3. Related Work
:Machine comprehension, Visual question answering

4. Question answering experiments
: using SQuAD(=a machine comprehension dataset on a large set of Wikipedia articles, with more than 100,000 questions.)
: During training, the moving averages of all weights of the model are maintained with the exponential decay rate of 0.999.
: At test time, the moving averages instead of the raw weights are used.

- Results: BIDAF achieves an EM score of 73.3 and an F1 score of 81.1, outperforming all previous approaches.
- Ablations: To evaluate the attention flow, we study a dynamic attention model, where the attention is dynamically computed within the modeling layer's LSTM, following previous work. -> This is in contrast with our approach, where the attention is pre-computed before flowing to the modeling layer.

5. Cloze test experiments
: Dataset - In a cloze test, the reader is asked to fill in words that have been removed from a passage, for measuring one's ability to comprehend text.
: Model Details - The model architecture used for this task is very similar to that for SQuAD with only a few small changes to adapt it to the cloze test.
: Results - BIDAF outperforms previous single-run models on both datasets for both val and test data. On the DailyMail test, our single-run model even outperforms the vest ensemble method.