Assignment #2 - Part II (C)

Constrained Beam Search

Team ACDC
14th April 2024

To add constraint to the beam search so that each output sequence contains at least 1 English token, we add constraint to the beam at each step.

The Constrained Beam

The elements of the beam can be thought of as sequences till now. When calculating candidate tokens for the next beam, we can add a token to each of the sequences in previous beam and update the log probabilities of the new sequence accordingly. The beam always has two parts:

- **Unconstrained Part (UC)**: The sequences in this part of beam may or may not have an English token.
- **Constrained Part (C)**: The sequences in this part of beam must have at least one English token.
- The lengths of UC and C are hyper-parameters satisfying $uc\_length + c\_length = beam\_size$.

Step in Constrained Beam

For determining next beam using current beam, the two parts of the next beam should be computed separately. Lets assume that the beam at the current step is B_0 and at the next step is B_1, then

- **Unconstrained Part**: The UC of B_1 is determined by taking the top-k next tokens of all of the sequences in B_0 and then picking top uc_length sequences out of these.
- **Constrained Part**: The C of B_1 is determined by taking the top-k tokens of sequences in C of B_0 and top-k English tokens of sequences in UC of B_1. Explanation is as follows:
  1. The sequences in C of B_0 already have an English token present, any token can be appended to these sequences to get a candidate sequence for C of B_1.
  2. The sequences in UC of B_0 may not have an English token, hence only English tokens may be appended at the end of these sequences to get candidate sequences for C of B_1.