Vocabulary-constrained Question Generation with Rare Word Masking and Dual Attention

Emil Biju
Indian Institute of Technology Madras

CODS-COMAD 2021

Question Generation

Question generation is the task of generating questions from a text passage that can be answered using information available in the passage.

Existing Work:

- Rule-based Approaches (Semantic Role Labelling, Syntactic Parsing, POS tagging)
- LSTM/GRU based Neural Encoder-Decoder frameworks
- Transformer/BERT based Question Generation

Applications:

- Automated tutoring and interviewing systems
- Chatbots and virtual assistants

Motivation

- Known models for Question Generation use a **large**, **predefined** vocabulary.
- Large Vocabulary: Increases memory usage, number of network parameters, computational requirements, training and inference times
- Predefined Vocabulary: May not include context-specific words from the input passage
- Studies show that 74-94% of conversations and written text in English are covered by the 1000 most common English words.

Motivation

How would a human construct a question?

Apollo 11 was the spaceflight that first landed humans on the Moon

1. Identify a keyword or phrase that the question can be based on

Apollo 11 was the spaceflight that first landed humans on the Moon

2. Choose an interrogative word (when, what, who, why, where or how)

3. Guess an outline for the question (prepositions [of, at, during, on]; articles [is, are, the]; conjunctions [that, and, or]; common words; punctuation marks)

what is the name of the ...... that ...... on the ......?

Motivation

what is the name of the ...... that ...... on the ......?

4. Point out other keywords from the statement to make the question more specific

Apollo 11 was the spaceflight that first landed humans on the Moon

5. Place the keywords at the right locations in the outline

what is the name of the spaceflight that first landed humans on the Moon ?

Learnings

1. Identify a keyword or phrase that the question can be based on

2. Choose an interrogative word (when, what, who, why, where or how)

3. Guess an outline for the question (prepositions [of, at, during, on]; articles [is, are, the]; conjunctions [that, and, or]; common words; punctuation marks)

Uses common words that can be derived from a limited-size vocabulary

4. Point out other keywords from the statement to make the question specific

5. Place the keywords at the right locations in the outline

Requires specific, uncommon words that can be derived from the input statement

Proposed Architecture

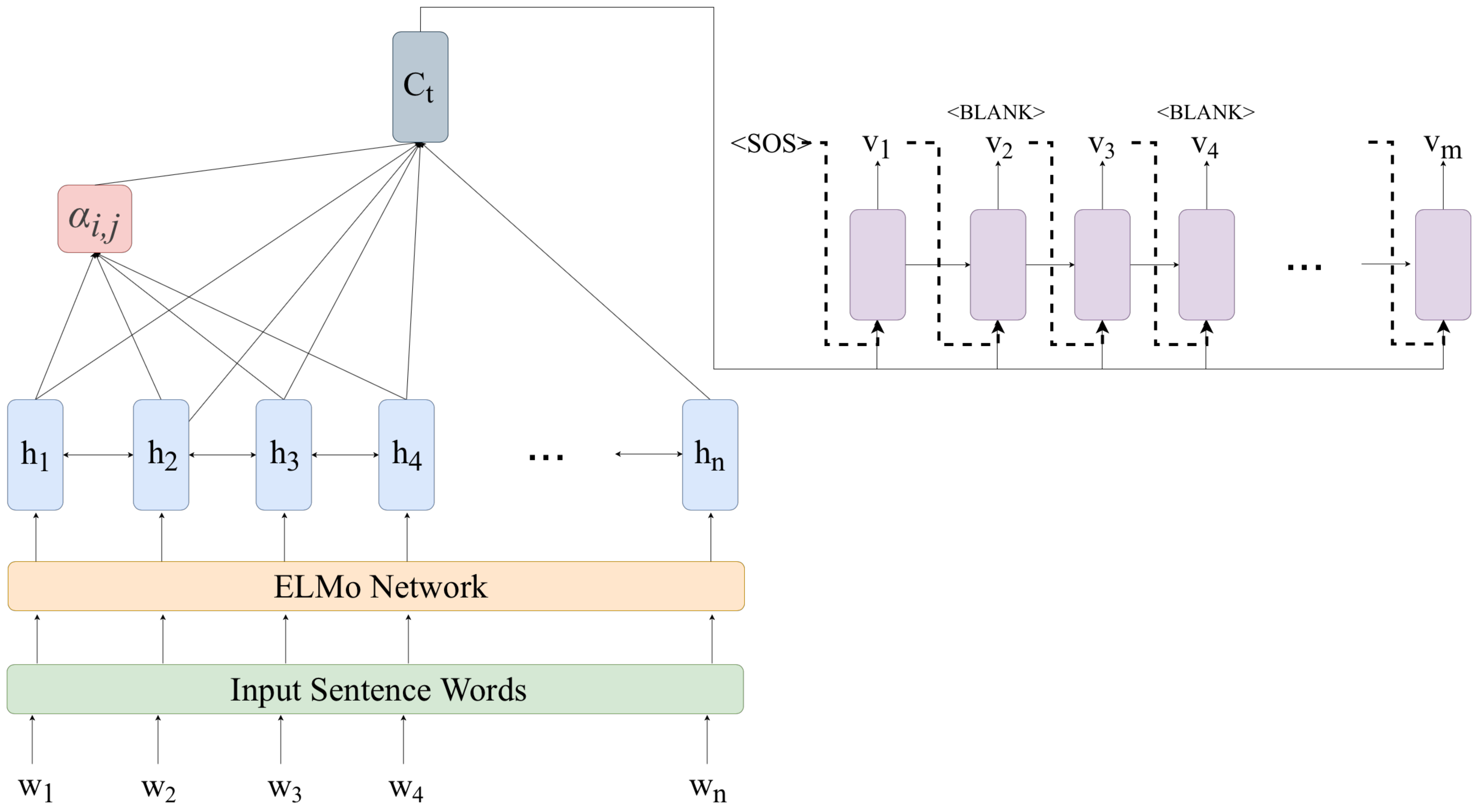
We propose to break the question generation task into two subtasks and develop an architecture composed of two neural network modules:

- **Question Structuring Module**
  - To generate the skeletal structure of a question using common words from the vocabulary and the structure of the input statement.
- **Word Pointer Module**
  - To point to rare words from the input passage to complete the question.
- For this work, we use question-answer pairs from the Stanford Question Answering Dataset (SQuAD)
- **Vocabulary:** We define the model vocabulary as the set of 1000 most common words from the SQuAD dataset

Question Structuring Module

The Question Structuring module creates the skeletal structure of a question.

- **Encoder:** A Bidirectional LSTM layer encodes the word embeddings into a sequence of vectors.
- **Decoder:** A Unidirectional LSTM layer either predicts a common word from the vocabulary or predicts the `<BLANK>` tag.
- Bahdanau Attention is applied over the encoder outputs to obtain a context vector that is passed to the decoder along with the previous predicted word



Word Pointer Module

How do we choose words that should replace the `<BLANK>` tags?

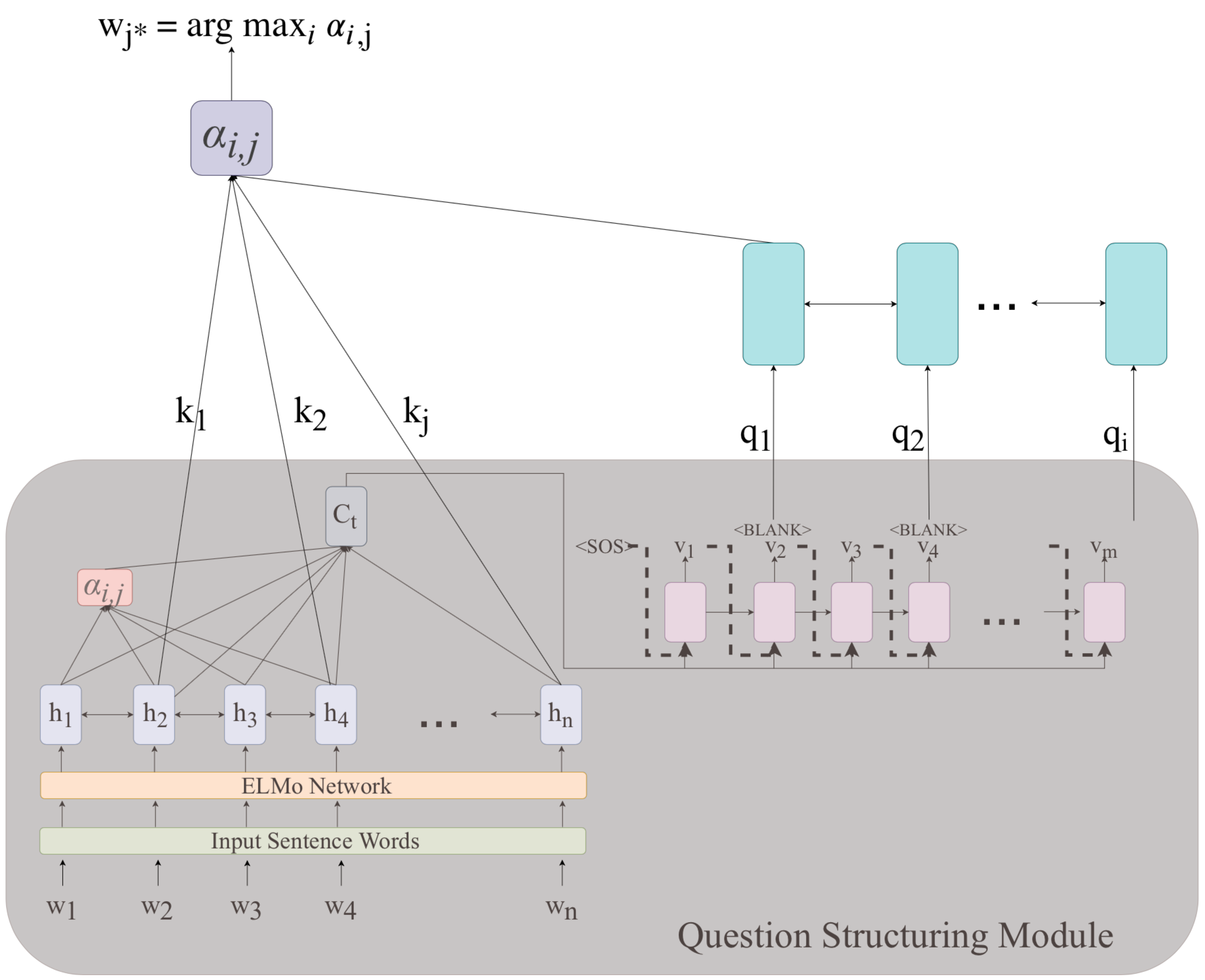
- Can we use the word that received the highest attention score during the decoding step that predicted the `<BLANK>` tag?
  - While this is a meaningful choice, attention scores are assigned by the model to help in the correct prediction of words from the common-words vocabulary. High attention scores may also be assigned to words that provide suitable context for decoding, apart from words that could fit in the decoded sentence.

- Can we train a separate attention network that uses the hidden states generated by the decoder to predict the `<BLANK>` tag?
  - Yes, these hidden states contain contextual information that required an uncommon word for expression. We can use the same information to point to the most relevant word.

Word Pointer Module

This module replaces the `<BLANK>` tags in *O* with context-specific words from the input sentence.

- We define $\{k_j\}$ as the set of encoder output vectors which correspond to words that are not from our vocabulary.
- We define $\{q_i\}$ as the set of vectors from the decoder output sequence which correspond to the prediction of the `<BLANK>` tag.
- We use an attention framework to compare each q_i with every k_j . The word that receives the highest attention score replaces the `<BLANK>` tag at that location.



An Example

- Consider the following input statement:

Columbus discovered America in 1492.

- The Question Structuring Module predicts common words from the vocabulary at appropriate locations to generate the skeletal structure of the question and predicts the `<BLANK>` tag otherwise.

When did <BLANK> discover <BLANK>?

- The Word Pointer Module predicts the most likely word from the input statement that should replace the `<BLANK>` tag at each location.

When did Columbus discover America?

Results and Future Work

We compared the performance of our model against an LSTM encoder-decoder framework trained on two different vocabulary sizes.

- The 1,000-word vocabulary used by our model: Our model achieves a higher BLEU score by 8.8%
- A 10,000-word vocabulary: Our model achieves nearly equal BLEU score compared to this variant

This technique can be extended to other similar tasks including translation, caption generation and summarisation to enable resource-constrained neural network models to train and perform efficiently.

Thank You!