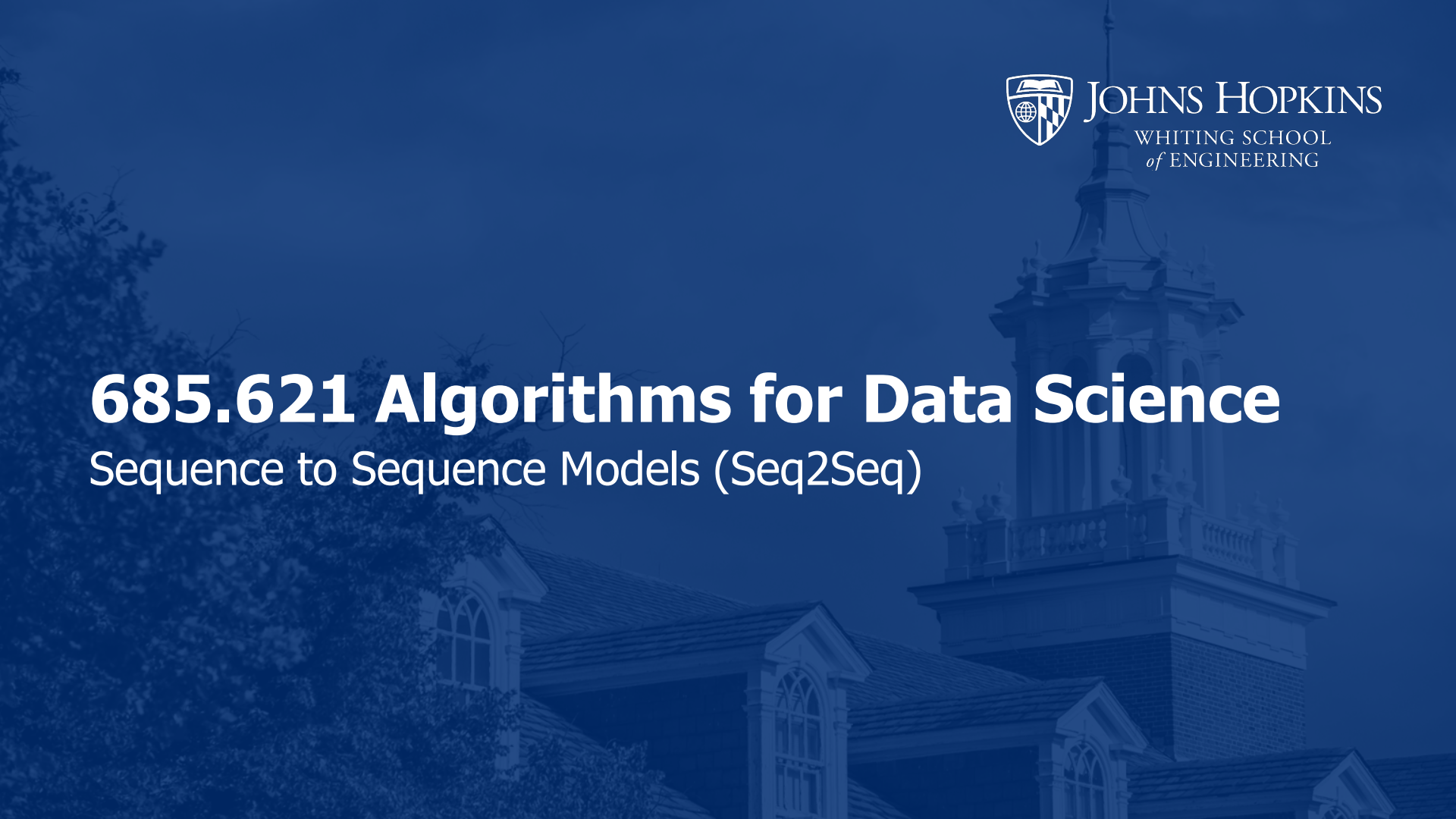


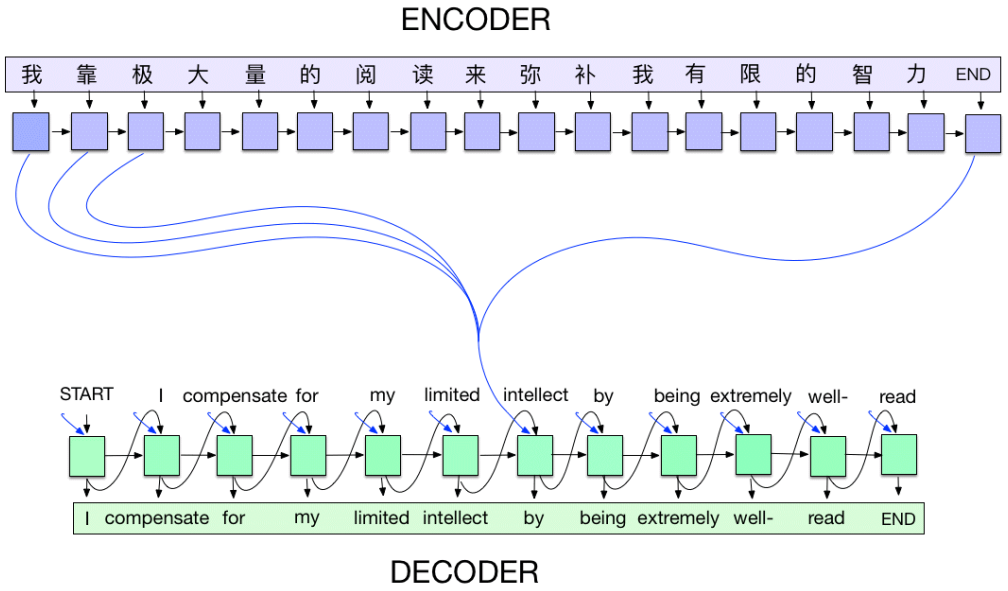
685.621 Algorithms for Data Science

Sequence to Sequence Models (Seq2Seq)

Introduction to Seq2Seqs

Seq2Seq models transform sequences from one domain into another using variable-length translation. It employes encoder-decoder architecture.

- **Basic Structure**
  - Encoder: Processes input sequence and compresses to fixed-size context vector
  - Decoder: Generates the output sequence using contect vector
  - Decoder: Reconstructs data from latent variables z.



Google OS Blog, 2021

Seq2Seq Objective

Seq2Seqs use maximum likelihood estimation as their loss function. Specifically, categorical cross-entropy loss is used to ensure the predicted sequence is as close as possible to the ground truth.

$$\mathcal{L}\text{Seq2Seq} = -\sum t = 1^{T} \log p(y_t | y_{<t}, \mathbf{x})$$

- Thus as each time step, we compare the model's predicted output with the actual output

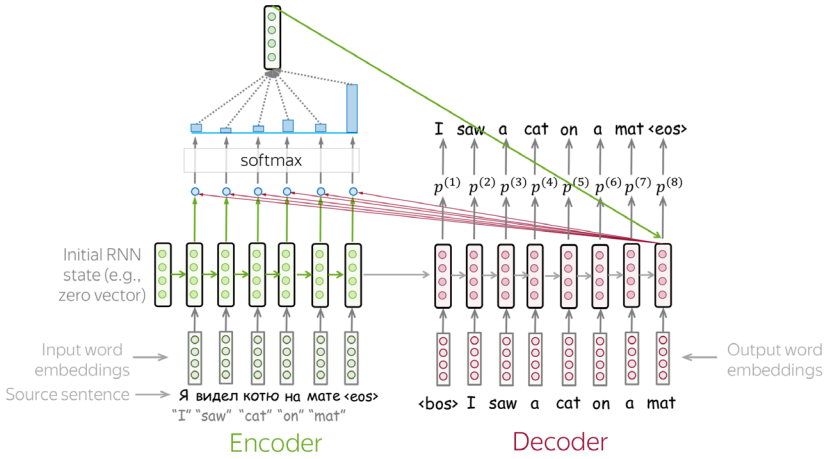
Step 1: Define the Encoder and Decoder

The Encoder processes the input and produces a context vector. The Decoder takes this context vector and generates output step-by-step. Utilizing attention helps model focus on relevant words.

- **Example of Encoder**
  - Begin with an input sequence which will be converted into vector representations through an embedding layer. Then use an LSTM encoder which encodes input sentence into hidden states.
- **Example of Decoder**
  - Use another LSTM to decode from the hidden states. Add an additional attention layer which allows decoder to focus on different encoder states at each step. Combine these two and add an activation layer to generate final words.

Step 2: Training/Evaluating the Seq2Seq

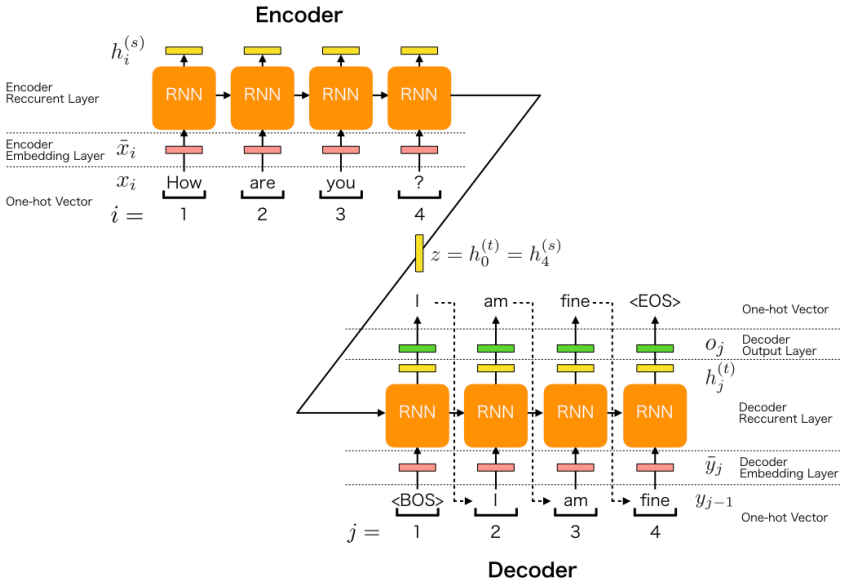
1. **Building**: Pass equivalent input and output sequences into encoder and decoder and build LSTMs for both.
2. **Training**: Initialize and train model on encoder inputs and decoder inputs so the model learns to translate with Attention helping the decoder focus on relevant words.
3. **Functionalize**: Build a function which creates tokenized sequences and utilizes the trained Seq2Seq model to translate.



Voita, 2023

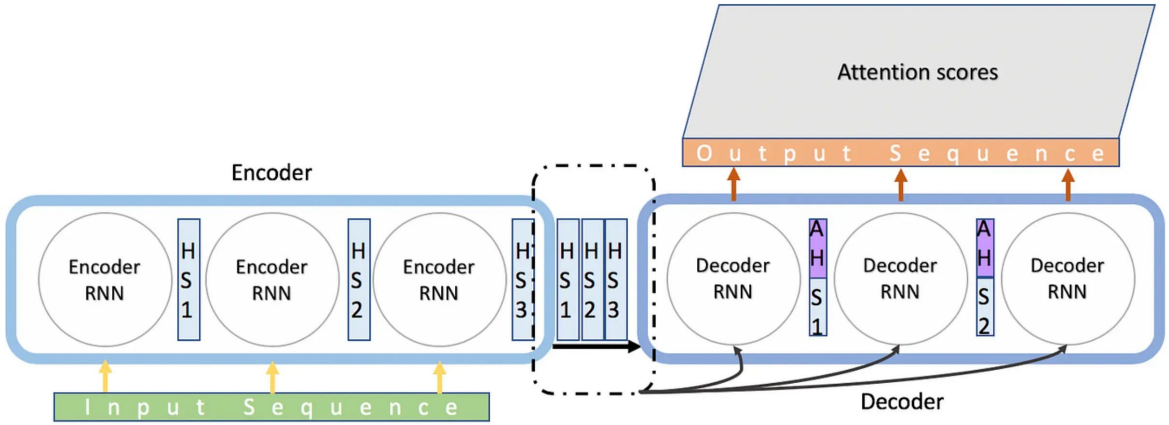
Step 3: Advantages

1. **Flexible Input/Output Lengths**: Can work with variable-length sequences and perform translation, summarization, and dialogue.
2. **Improves with Attention Mechanism**: Performance is improved for long sequences through focus on relevant parts
3. **Handles Dependencies**: Captures long term dependencies through use of LSTMs



Step 4: Challenges

1. **Memory Intensive**: Long sequences require many hidden states.
2. **Exposure Bias**: Model uses its own outputs for inference which could cause errors.
3. **Degrades without Attention**: Performance is degraded for long sequences



Step 5: Real World Applications

1. **Speech-to-Text**: Converts spoken language into text.

2. **Image Captioning**: Generates a descriptive sentence from an image.

3. **Subtitle Generation**: Utilize audio waveforms and speech recognition models combined with Seq2Seq to generate subtitles.