Name: Suhaila Ahmed Hassan
Track: AI
Branch: Alexandria

NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE

Introduction

Neural Machine Translation (NMT) aims to build a single, end-to-end neural network for translation, as opposed to traditional phrase-based systems made of many components.

Most NMT models use an encoder-decoder architecture where:

- The encoder converts the source sentence into a fixed-length vector.
- The decoder generates the translation from this vector.

Problem with Approach:

- It struggles to handle long sentences, especially those longer than seen during training.
- Performance deteriorates as sentence length increases

Proposed Model

Neural machine translation architecture combining:

- A bidirectional RNN encoder.
- A decoder that emulates searching through a source sentence during decoding a translation.

Decoder:

- Decoder in this model uses a different context vector for each output word.
- Context vector: Computed as a weighted sum of annotations from the encoder, using attention weights that indicate relevance of each source word to current target word.
- Attention scores: The weights are calculated using an alignment model, a small neural network trained jointly with the main model.
- Soft alignment (attention): Instead of selecting a single aligned word, the model computes a weighted average across all source words. This soft alignment allows for differentiable training via backpropagation.

Encoder:

- Uses a bidirectional RNN (BiRNN) where one RNN reads the sequence from left to right and another one reads it from right to left.
- To obtain annotations for each word, the model concatenates the forward and backward hidden states. These annotations capture context from both directions for past and future.

Experiment

- The model is evaluated on English-to-French translation using the WMT 2014 dataset.

- Training is done using only parallel corpora (no monolingual data).
- Models are trained on sentence lengths up to 30 and 50 words, and evaluated using BLEU scores.
- Models to compare: traditional RNN encoder-decoder and proposed RNNsearch.

Results

Proposed RNNsearch model significantly outperforms the traditional RNN Encoder Decoder. It achieves higher BLEU scores and shows better performance with longer sentences.