NLP Assignment 3

Backpropagation through Time (BPTT)

Backpropagation through Time (BPTT) is a variant of the backpropagation algorithm used to train Recurrent Neural Networks (RNNs). BPTT extends the standard backpropagation algorithm to handle sequences of data by unfolding the network through time. During training, BPTT iteratively updates the network's weights by computing gradients of the loss function with respect to the network parameters at each time step and propagating them backward through time. BPTT is used to optimize the RNN's parameters to minimize the prediction error over the entire sequence.

Vanishing and Exploding Gradients

Vanishing and exploding gradients are common issues that occur during the training of deep neural networks, including RNNs. Vanishing gradients occur when the gradients of the loss function with respect to the network parameters become very small, causing the network to learn slowly or not at all. Exploding gradients occur when the gradients become very large, leading to unstable training and divergent behavior. These issues are particularly problematic in RNNs due to the recurrent nature of the network, where gradients can accumulate or diminish over many time steps.

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of RNN architecture designed to address the vanishing gradient problem and capture long-range dependencies in sequential data. LSTM cells include additional memory units called cells and a set of gates (input gate, forget gate, and output gate) that regulate the flow of information. The gates control which information is stored in the memory, which information is discarded, and which information is passed to the output. LSTM cells can effectively learn and remember patterns in sequential data over long time scales.

Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) is another type of RNN architecture that simplifies the design of LSTM cells while maintaining similar performance. GRU cells also include gating mechanisms to control the flow of information, but they combine the input and forget gates into a single "update gate" and use a different mechanism to update the cell state. GRU cells have fewer parameters and are computationally more efficient than LSTM cells, making them popular for applications with limited computational resources.

Peephole LSTM

Peephole LSTM is an extension of the standard LSTM architecture that includes additional connections (peepholes) between the cell state and the gates. These connections allow the gates to directly access the current cell state, providing additional information to make more informed decisions. Peephole connections enable LSTM cells to better capture temporal dependencies and improve performance on tasks requiring precise timing or long-term memory.

Bidirectional RNNs

Bidirectional Recurrent Neural Networks (RNNs) are RNN architectures that process input sequences in both forward and backward directions. Bidirectional RNNs consist of two separate RNN layers, one processing the input sequence in the forward direction and the other processing it in the backward direction. The outputs of both layers are concatenated or combined in some way to produce the final output sequence. Bidirectional RNNs can capture information from both past and future contexts, making them effective for tasks where context from both directions is important, such as sequence labeling and sequence-to-sequence modeling.

BiLSTM

Bidirectional Long Short-Term Memory (BiLSTM) is a variant of the LSTM architecture that combines forward and backward LSTM layers to capture information from both past and future contexts. In a BiLSTM, the input sequence is processed by two separate LSTM layers: one layer processes the input sequence in the forward direction, and the other layer processes it in the backward direction. The outputs of both layers are typically concatenated or combined in some way to produce the final output sequence. BiLSTMs are commonly used in tasks such as sequence labeling, sentiment analysis, and machine translation, where capturing bidirectional context is crucial for achieving high performance.

BiGRU

Bidirectional Gated Recurrent Unit (BiGRU) is similar to BiLSTM but uses GRU cells instead of LSTM cells for processing the input sequence in both forward and backward directions. Like BiLSTMs, BiGRUs capture bidirectional context by processing the input sequence from both past and future contexts. BiGRUs have fewer parameters than BiLSTMs and are computationally more efficient, making them suitable for applications with limited computational resources. BiGRUs are widely used in various sequence modeling tasks, including text classification, named entity recognition, and speech recognition.