NLP assignment-5

1. Sequence-to-sequence models are a type of neural network architecture that are commonly used for natural language processing tasks such as machine translation, summarization, and text generation. They consist of two recurrent neural networks (RNNs), one for encoding the input sequence and one for generating the output sequence.
2. One of the main problems with vanilla RNNs is that they have difficulty learning long-term dependencies. This means that they are not able to effectively process input sequences that have a large time lag between important events or have a long-term context.
3. Gradient clipping is a method used to prevent the gradients of a neural network from becoming too large, which can cause problems during training. It involves scaling down the gradients to a maximum value if they exceed a certain threshold.
4. Attention mechanisms allow a neural network to focus on certain parts of an input sequence when processing it. They work by weighting the input sequence so that the network pays more attention to certain parts of the sequence and less to others. This allows the network to selectively focus on the most relevant information when processing the input.
5. Conditional random fields (CRFs) are a type of probabilistic model used for sequence labeling tasks such as part-of-speech tagging and named entity recognition. They are used to model the conditional probability of a sequence of labels given a sequence of input observations.
6. Self-attention is a type of attention mechanism that allows a neural network to attend to different positions of the input sequence in order to compute a weighted sum of the inputs. This is useful for tasks where the order of the input sequence is important, such as language translation.
7. Bahdanau attention is a type of attention mechanism that was introduced in a 2014 paper by Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. It works by using a small neural network to compute the attention weights for each position in the input sequence.
8. A language model is a statistical model that is used to predict the likelihood of a sequence of words in a language. It can be used for tasks such as speech recognition, machine translation, and text generation.
9. Multi-head attention is a variant of the attention mechanism that allows a neural network to attend to multiple parts of the input sequence simultaneously. It works by splitting the attention mechanism into multiple "heads," each of which attends to a different part of the input sequence and then concatenating the results.
10. Bilingual Evaluation Understudy (BLEU) is a metric for evaluating the quality of machine translation systems. It works by comparing the machine-translated output to a reference translation and calculating the degree of overlap between the two. The higher the overlap, the higher the BLEU score.

Top of Form