gated units (LSTM and GRU) significantly outperform traditional tanh-RNNs, which struggle with vanishing gradients and fail to model long sequences effectively. While both LSTM and GRU demonstrate strong performance, GRUs sometimes converge faster and achieve comparable or slightly better results than LSTMs, depending on the dataset. However, no definitive winner emerges between the two, as their effectiveness varies by task.

A key advantage of gated units is their additive update mechanism, which helps preserve long-term information and improves gradient flow during training. LSTMs offer finer control over memory exposure through an output gate, whereas GRUs are simpler with fewer gates but lack explicit memory regulation.

The study concludes that gated RNNs (LSTM and GRU) are superior to traditional tanh-based RNNs for sequence modeling tasks. The choice between LSTM and GRU may depend on the specific application, with GRUs offering potential efficiency benefits and LSTMs providing more precise memory management. These findings support the use of advanced gated architectures in applications like speech recognition, music generation, and machine translation.

|  |
| --- |
| The attention is a general mechanism in neural networks that allows models to focus on specific parts of the input when generating output. It helps with long sequences by computing a weighted sum of encoder states, where the weights (attention scores) indicate relevance to the current decoding step. |

This neural machine translate by the attention mechanism, which has since become a cornerstone of modern sequence-to-sequence models. The approach addresses the information bottleneck of fixed-length representations while providing interpretable insights into the translation process through its learned attention patterns. The success of this architecture paved the way for subsequent developments in attention-based models, including the Transformer architecture that now dominates the field.

The loung attention is a specific type of attention mechanism introduced by Minh-Thang Luong et al It proposed global and local attention mechanisms, and different ways of calculating alignment scores dot, concat. Attention mechanisms for neural machine translation, introducing and comparing two fundamental approaches: global and local attention. The global attention model maintains the conventional approach of attending to all source words, while the local attention model innovatively focuses on only a subset of source positions at each time step, offering computational efficiency, Key technical contributions include the development of different alignment scoring functions and an input-feeding approach that maintains awareness of past alignment decisions. The analysis reveals important insights about model behavior, particularly in handling long sentences and producing accurate word alignments. The local attention mechanism emerges as particularly valuable, combining the benefits of both soft and hard attention approaches while remaining computationally tractable.

ROUGE, or Recall-Oriented Understudy for Gisting Evaluation, is a set of metrics and a software package used for evaluating automatic summarization and machine translation software in natural language processing. The metrics compare an automatically produced summary or translation against a reference or a set of references (human-produced) summary or translation. ROUGE metrics range between 0 and 1, with higher scores indicating higher similarity between the automatically produced summary and the reference.