GRU

GRU: Gated recurrent unit

GRUs are an innovative variation of Recurrent Neural Networks, designed to improve upon and simplify the architecture of LSTMs. They offer a more streamlined approach to handling sequential data, particularly effective in scenarios where long-term dependencies are vital.

How GRU Works:

- **Simplified Architecture:** The GRU is known for its simplified structure compared to LSTM, making it more efficient in terms of computational resources. This efficiency stems from its reduced number of gates.
- **Gating Mechanism:** GRUs utilize two gates:

> Update Gate: This gate decides the extent to which the information from the previous state should be carried over to the current state. It is a blend of the forget and input gates found in LSTMs.
>
> Reset Gate: It determines how much of the past information to forget, effectively allowing the model to decide how much of the past information is relevant for the current prediction.

No Separate Cell State: Unlike LSTMs, GRUs do not have a separate cell state. They combine the cell state and hidden state into a single structure, simplifying the information flow and making them easier to model and train.

Challenges and Advantages:

- GRUs are particularly known for their efficiency and speed in training, making them a suitable choice for models where computational resources are a concern.
- While they are generally faster and simpler than LSTMs, they might not be as effective in capturing very long-term dependencies due to their simplified structure.

Fonte

Neural Networks in NLP: RNN, LSTM, and GRU
https://medium.com/@mervebdurna/nlp-with-deep-learning-neural-networks-rnns-lstms-and-gru-3de7289bb4f8