

Ans to the Q/A no. 3

GRU or Gated Recurrent Unit aims to solve the vanishing gradient problem which comes with a standard ~~recu~~ neural network. To solve the vanishing gradient problem of standard GRU, we use update gate and reset gate. Basically, these are two vectors which decide what information should be passed to the output. The special thing about them is that they can be trained to keep information from long ago, without washing it through time or remove information which is irrelevant to the prediction. The update gate helps the model to ~~det~~ determine how much of the past information (from previous time steps) needs to be passed along to the future. The reset gate is used from the model to decide how much of the past information to forget.

The main difference of LSTM and GRU is in the Gate number. LSTM has three gates (input, output and forget) whereas GRU has two gates (reset and update gate). GRU couples forget as well as input gates. GRU use less training parameters and therefore use less memory, execute faster and train faster than LSTM's whereas LSTM is more accurate on dataset using longer sequence. In short, if sequence is large or accuracy is very critical, ~~please~~ we should go for LSTM whereas for less memory consumption and faster operation go for GRU. If we don't have much floating point operations per second (FLOPs) to spare which to GRU. LSTM has three values at output (output, hidden and cell) whereas GRU has two values at output.

GRU has two gates and faster than compute to RNN. And LSTM networks are an extension of RNN that extend the memory. GRU is faster because GRU ~~use~~ use less training parameters and therefore use less memory, execute faster and train faster than ~~GRU~~ RNN. RNN is more accurate on datasets using longer sequence. In short, if sequence is large or accuracy is very critical.