GRU
2 gates
no external memory
less training param
less memory
faster.

LSTM
3 gates

computationally faster ?

M23CS7.501:
Advanced Natural Language Processing

Mid Semester Exam

Sep 22, 2023

only one ↗ 3

MM: 30 Time: 1.5 hrs

Note: Marks are mentioned next to the questions.

1. Explain how RNNs enable handling long distance dependencies. Can you compare LSTMs and GRUs at an architectural level?
[5+5 marks]
2. Describe the ELMO Contextual embedding model. How is it used for downstream tasks?
[5+5 marks]
3. Describe a sequence-to-sequence machine translation model where decoder pays attention to the encoder. Why is the attention needed?
[3 + 7 marks]

1) - Recurrance also talk about Fwfo bapogation
- Name of gates is not imp. Similarity

2) ELMO - standard Bi LSTM
- weighted summation of Layer emb.
- weights to be computed based on task

3) Description RNN/LSTM with att 'fwd' depend
- Diff Target words need diff context words
for Translation.