



Method	Advantages	Disadvantages
LSTM	<ul style="list-style-type: none"> - models long-term dependencies better than a simple RNN - more robust to vanishing gradients than a simple RNN 	<ul style="list-style-type: none"> - higher memory requirement and computational complexity than a simple RNN due to multiple memory cells
S-LSTM	<ul style="list-style-type: none"> - models complicated inputs better than LSTM 	<ul style="list-style-type: none"> - higher computational complexity in comparison with LSTM
Stacked LSTM	<ul style="list-style-type: none"> - models long-term sequential dependencies due to deeper architecture 	<ul style="list-style-type: none"> - higher memory requirement and computational complexity than LSTM due to a stack of LSTM cells
Bidirectional LSTM	<ul style="list-style-type: none"> - captures both future and past context of the input sequence better than LSTM and S-LSTM 	<ul style="list-style-type: none"> - increases computational complexity in comparison with LSTM due to the forward and backward learning
Multidimensional LSTM	<ul style="list-style-type: none"> - models multidimensional sequences 	<ul style="list-style-type: none"> - higher memory requirement and computational complexity than LSTM due to multiple hidden state vectors - instability of the network as grid size and depth grows
Grid LSTM	<ul style="list-style-type: none"> - models multidimensional sequences with increased grid size 	<ul style="list-style-type: none"> - higher memory requirement and computational complexity than LSTM due to multiple recurrent connections
Differential RNN	<ul style="list-style-type: none"> - discrimination between salient and non-salient information in a sequence - better captures spatiotemporal patterns 	<ul style="list-style-type: none"> - increases computational complexity in comparison with LSTM due to the differential operators
Local-Global LSTM	<ul style="list-style-type: none"> - improves exploitation of local and global contextual information in a sequence 	<ul style="list-style-type: none"> - increases computational complexity in comparison with LSTM due to more number of parameters for local and global representations
Matching LSTM	<ul style="list-style-type: none"> - optimizes LSTM for natural language inference tasks 	<ul style="list-style-type: none"> - increases computational complexity due to word-by-word matching of hypothesis and premise
Frequency-Time LSTM	<ul style="list-style-type: none"> - models both time and frequency 	<ul style="list-style-type: none"> - more computational complexity than LSTM due to more number of parameters to model time and frequency

Long-Short Term Memory(LSTM)

Standard LSTM

S-LSTM

Stacked LSTM

Bidirectional LSTM

Multidimensional LSTM

Grid LSTM

Differential Recurrent Neural Networks

Other LSTM Models