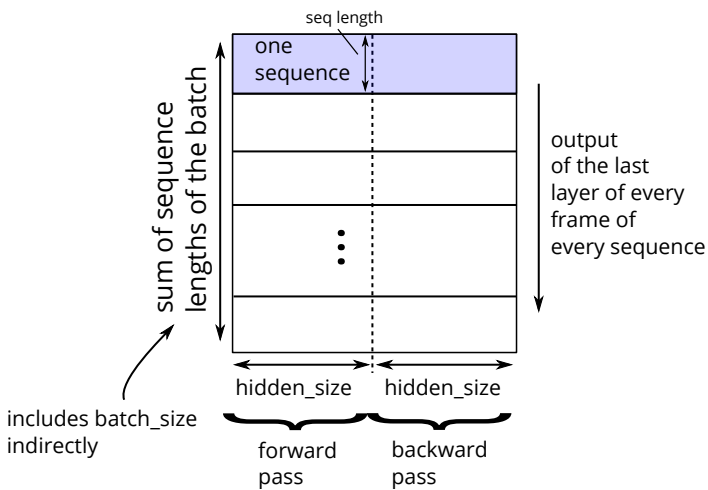


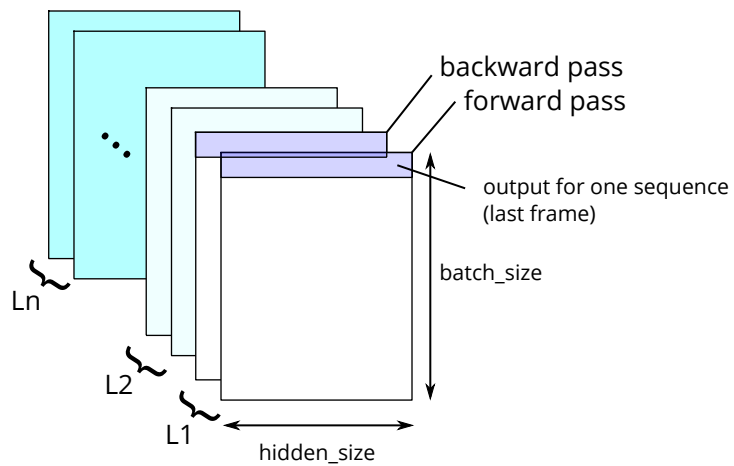
lstm output (w/ packed_sequence input):



output = [sum of seq. lengths, D*hidden_size]

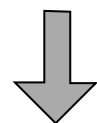
D = 1 (unidirec. lstm) or 2 (bilstm)

h_T output:



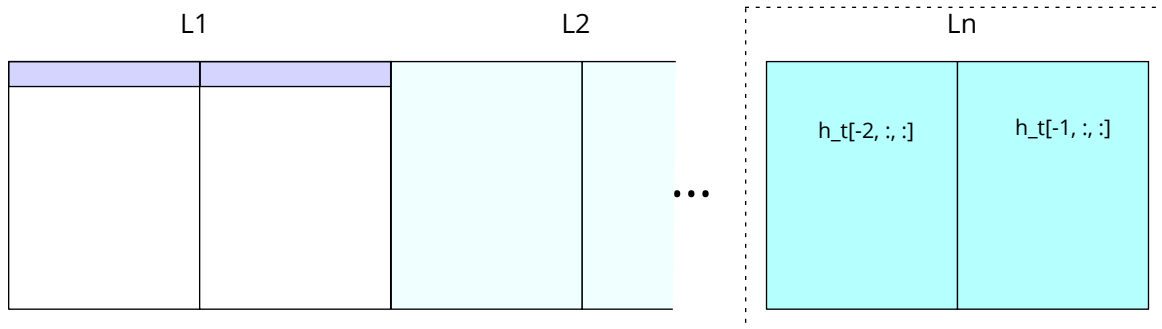
L = Layer

$h_T = [D \times L, \text{batch_size}, \text{hidden_size}]$



Reshaping for follow-up linear layer

reshaped to [batch_size, D*L*hidden_size]:



relevant output

$h_T = \text{cat}([h[-1, :, :], h[-2, :, :]], \text{dim}=-1)$

(reshaping not really needed, simply use the last (for lstms) or last two (bilstms) layer's output as shown in blue)