



I am trying to implement a Chatbot using Tensorflow and its implementation of seq2seq.

After reading different tutorials ([Chatbots with Seq2Seq](#), [Neural Machine Translation \(seq2seq\) Tutorial](#), [Unsupervised Deep Learning for Vertical Conversational Chatbots](#)), and the original paper [Sequence to Sequence Learning with Neural Networks](#), I could not find an explanation as to why the Tensorflow seq2seq implementation **pads all sequences** (both input and output) to the **same fixed length**.

Example:

Input data consists of sequences of integers:

```
x = [[5, 7, 8], [6, 3], [3], [1]]
```

RNNs need a different layout. Sequences shorter than the longest one are padded with zeros towards the end. This layout is called `time-major`.

```
x is now array([[5, 6, 3, 1],  
               [7, 3, 0, 0],  
               [8, 0, 0, 0]])
```

Why is this padding required?

[Source](#) of this tutorial.

If I am missing something, please let me know.

tensorflow

chatbot

sequences

asked Jan 19 '18 at 11:02



[surveyCorps](#)

123 3 9

Short answer: for batching – [Maxim](#) Jan 19 '18 at 11:43

Thank you for your response. Could you please elaborate a bit more? – [surveyCorps](#) Jan 19 '18 at 11:57

1 Answer



You need to pad the sequence (with some id, in your case 0) to the maximum sequence length. The reason you want to do this is so that sequences can fit in an array (tensor) object and be processed in the same step.