Character-level Machine Translation

Neural Machine Translation

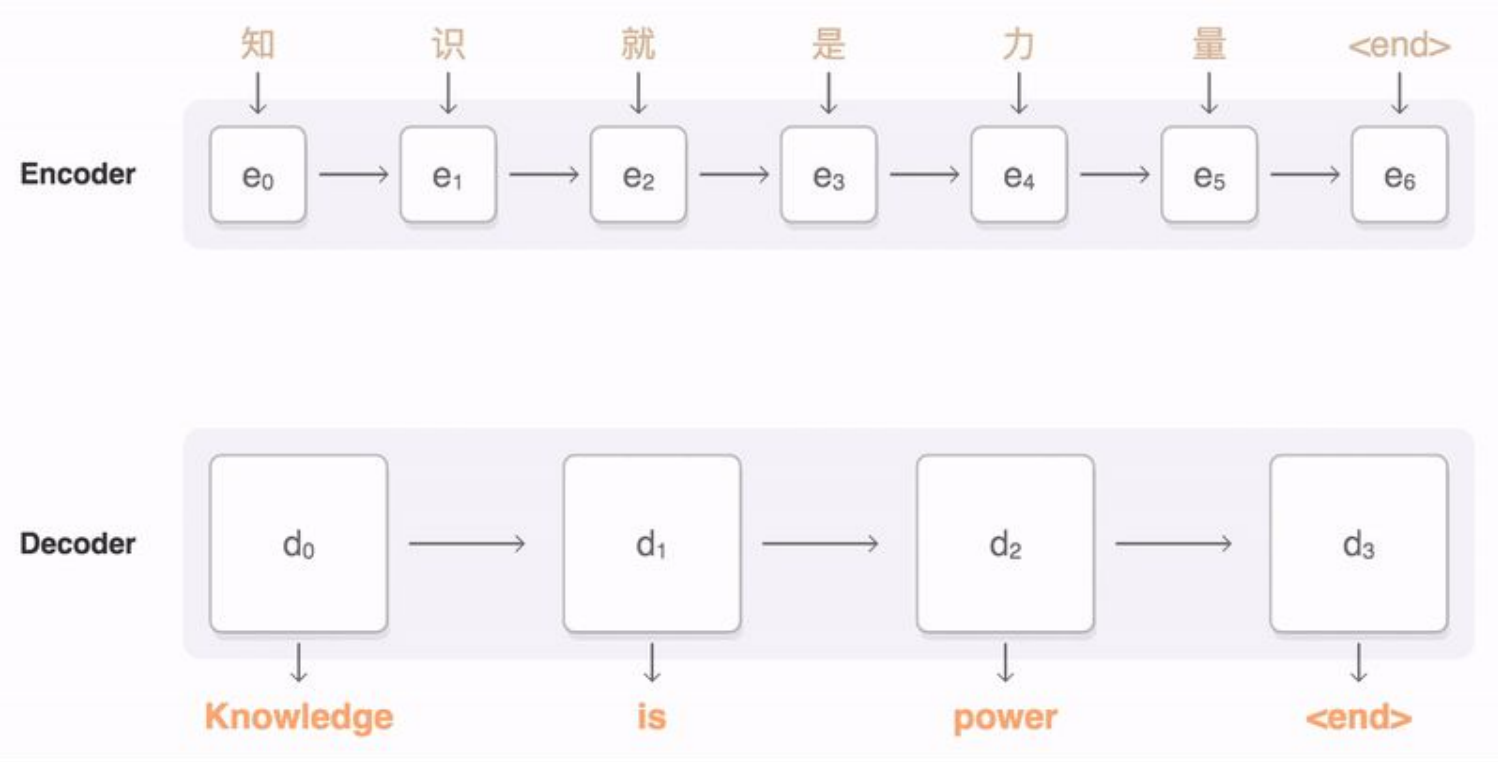
October, 2016

Alexander Rosenberg Johansen (github.com/alrojo)

Using Neural Networks to translate text

The computers are coming! → Les ordinateurs sont à venir!

Current method - Seq2Seq with attention on words



Google NMT - https://research.googleblog.com/2016/09/a-neural-network-for-machine.html

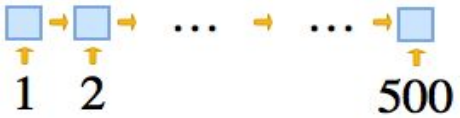
Contributions

New! TensorFlow Seq2Seq functions

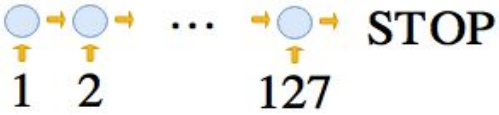
New! Hierarchical char-to-char NMT

TensorFlow Additions

Current RNN-Decoder

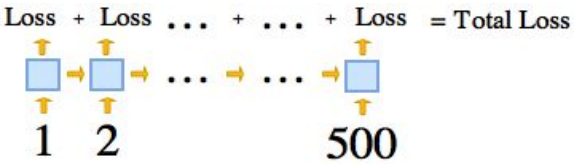


New! RNN-Decoder

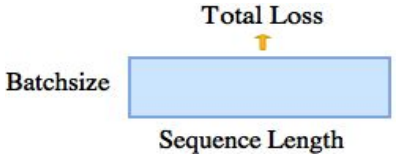


- \+ Variable Sequence Length
- \+ Masking Support

Current Sequence Loss



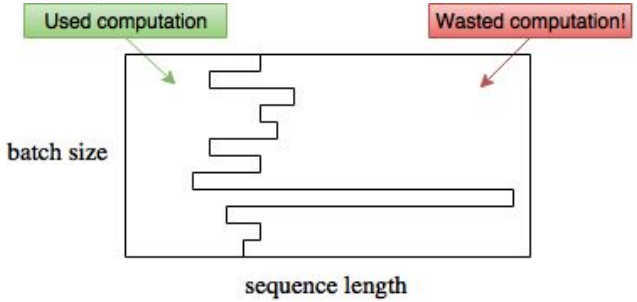
New! Sequence Loss



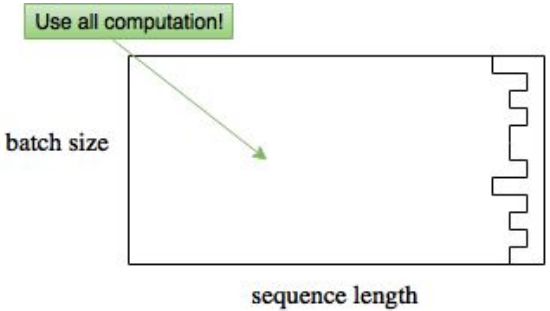
- \+ Faster
- \+ Variable Sequence Length

Dynamic batching

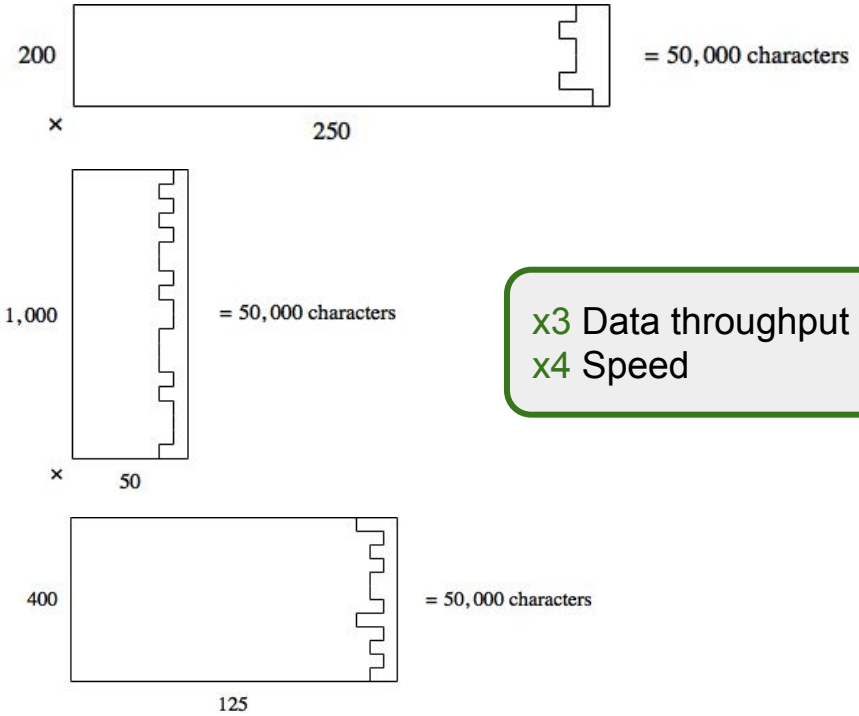
Regular batching



Bucketing

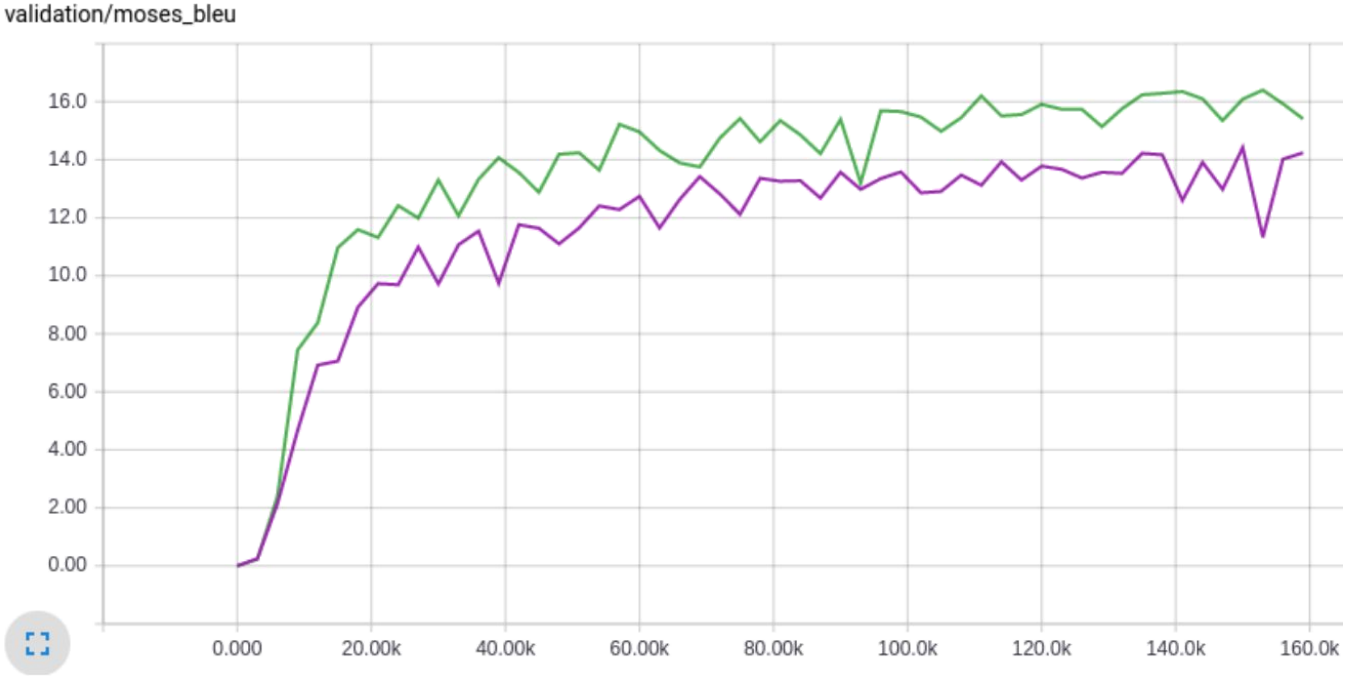


Dynamic batching = ***Bucketing*** *and* ***Variable Sequence Length***



x3 Data throughput
x4 Speed

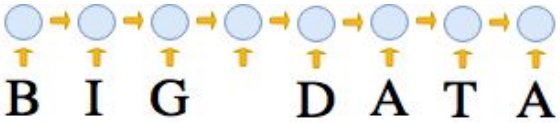
Serving Data - Dynamic Batching (same model)



Dynamic batching
Regular batching

Why words?

Chars VS Words



x5 Length

BIG DATA

$\{a, b, \ldots, z\}$

x250 Size

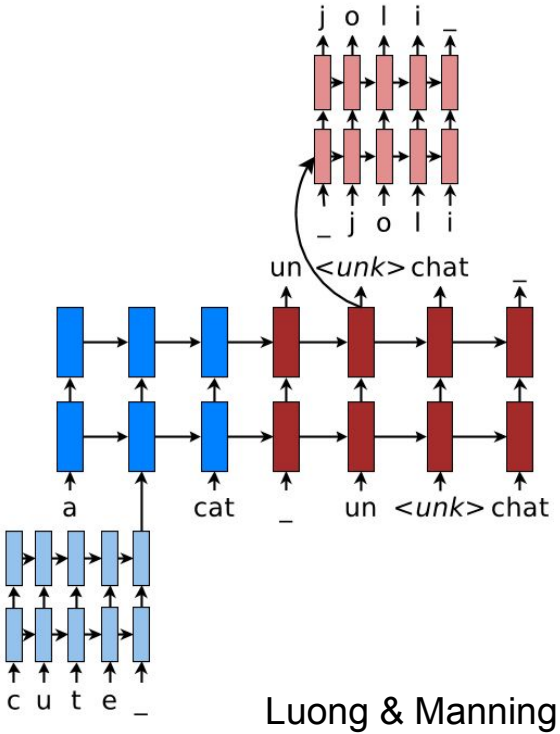
$\{abase, abbess, \ldots, zodiac\}$

$r\frac{a}{u}n$

Preserves relation!

$\frac{ran}{run}$

Most recent work: Word-Char hybrid

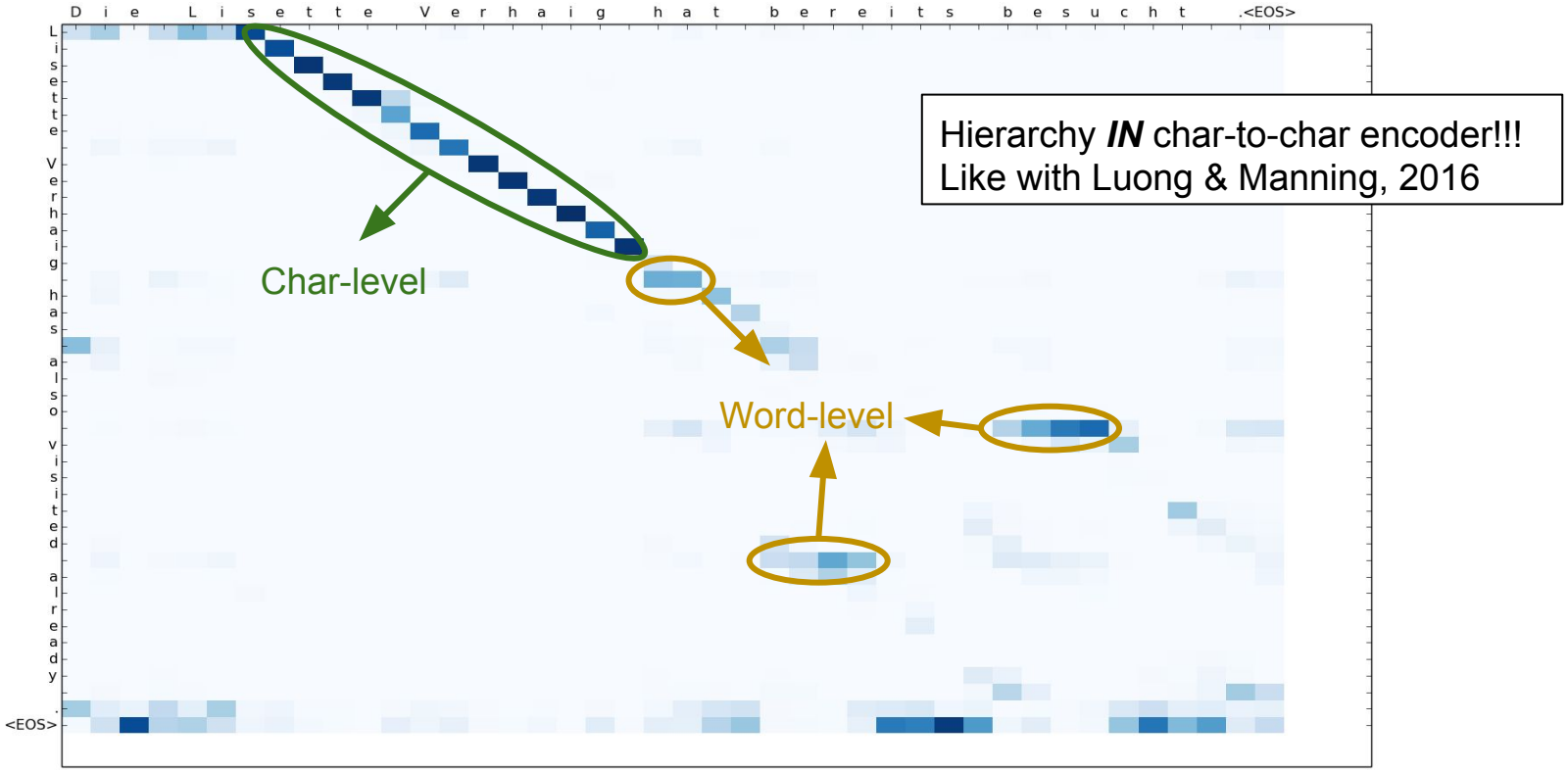


Luong & Manning, 2016

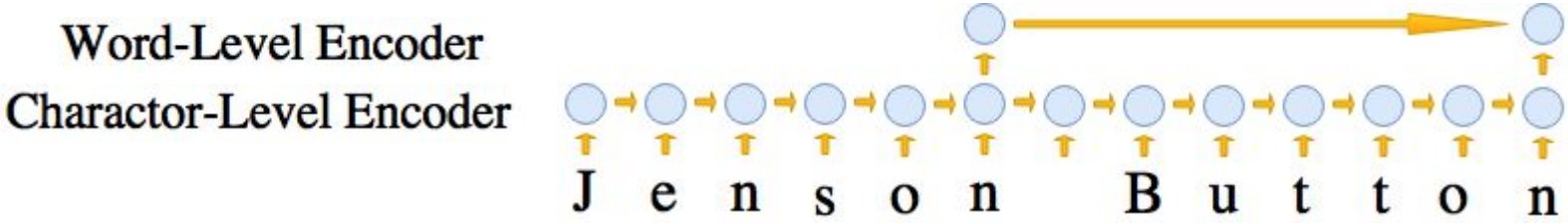
Use chars-to-char for out-of-dictionary

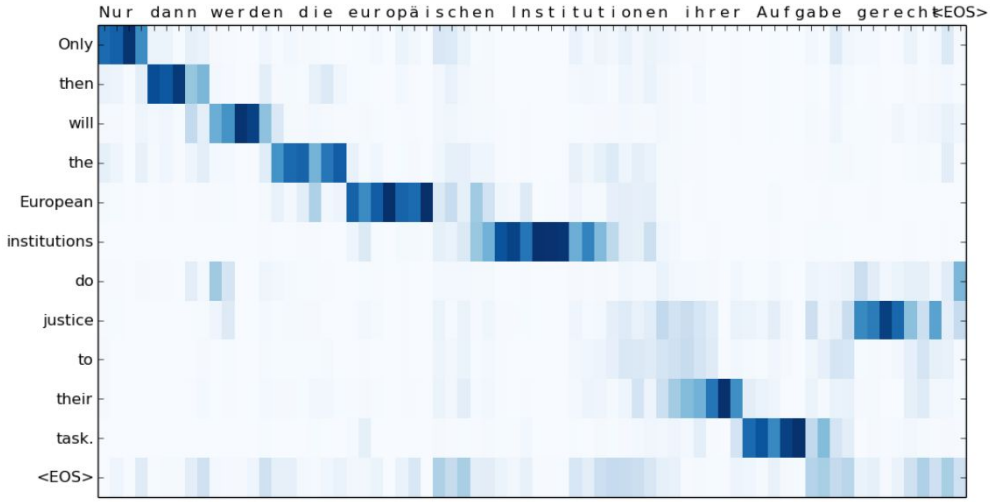
\+ Reduced dictionary size!

Char-to-Char attention plot (our model)



Hierarchical Encoding (our model)





Model	Language	newstest2015
`char-to-char`	En-De	15.14 BLEU
`char2word-to-char`	En-De	**17.43** BLEU