

new token.

Let's ~~know~~ preprocessing: text input Tokenizer

my dog my ~~the~~ word 那只狗
那只狗

oov_token = "< oov >"

pad-sequences

~~put~~ zeros
before each
sentence

Line:

Input Sequences

~~266~~ 8 67 68 69 70

[4 2]

[4 2 66]

[4 2 66 8]

[4 2 66 8 67]

[4 2 66 8 67 68]

pad 0

~~ex)~~ ~~[0 0 ... 0 4 2]~~ ~~Label(X)~~
[0 0 ... 0 4 2 66]

tf.keras.utils.to_categorical

(labels, num_classes = total_classes)

one-hot