

Oren Cohen
ID: 305164295
USER: cohenorx

- Can the two languages be distinguished using a bag-of-words approach?
No – one problem is that we get changed sized inputs – (we can solve that by padding)
another problem is that bag of words is not refers to the order of characters.
So the word: “1a2b3c4b” (which is good) will be the same as “1a2cbc4b” (which is bad)
- Can the two languages be distinguished using a bigram or trigram based approach?
Same problems here , for random sequence, all bigram or trigram can be labeled as 1, or 0
(12, aa, bb, 3b, cc... - its same for “good” examples and “bad” examples)
- Can the two languages be distinguished using a convolutional neural network?
Same as before, the main problem with CNN it’s not refers to the structure of the input.
And by definition, this problem differ only by ordering characters.