

a3-writeup

Guanxiong Liu

March 2018

Part 1.

1. It will perform very bad because for long sequences, the decoder will receive very little information of the first few character about the input.

2. roomba - omcerway
concert - orcortcay
hello - erlehay
table - adletay
translate - artartedray
possibility - orlerereyepay
interactive - irtiraariedway
information - inmonairotnway
we - eway
you - yoway
hi - ihay
ie - ieway
i - iway
g - uay
ca - apay
zh - utay
bow - owbay

The performance was very bad. However, if the input word has one or two characters, it relatively well. It has high possibility that predict the word correctly. If the word has more than two characters, it is most likely fail the prediction. Therefore, the longer the word, the lower the possibility that the translation is correct.

Part 2.

1. Teacher forcing might be limited when the generated characters vary from what was seen by the model during training. Since during training, we have access to ground truth, so we can force the model. During testing, we do not have access to ground truth, the output from previous time step becomes input. The generated sequences may diverge/become unstable.

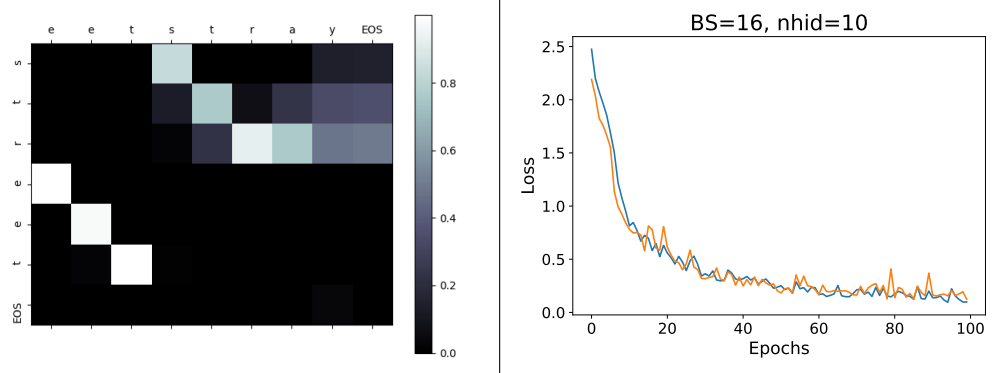
2. We can use decay probability to decide whether to use ground truth token. Since at first, the model is initialized randomly. We do not want the model to be trained on incorrect output of previous time step, thus more incorrect output should we get. Therefore, we have a decay probability that have model trained guided by teacher forcing. After we have more iterations, we should have a larger probability that use previous time step's output as input and gradually get rid of teacher forcing.

Part 3.

Please see the attached code.

Part 4.

Please see the attached code and the 99 epoch result, loss.



Part 5.

cake - akecay
make - akemay
even - evenway
drink - inkdray
cycle - yclecay
thesis - esisthay
aardvark - aardvarkway
oak - oakway
well-known - ellway-oinknnay
well-mannered - ellway-annereedpay
opt-out - optayay-outay
a - away
d - dayday
oooo - ooooway
asuperlongwordtested - asuperlongioresetete

{cake,make,even,drink,cycle,thesis,aardvark,oak,a,oooo} are translated correctly. {well-known,well-mannered,opt-out,d,asuperlongwordtested} are translated incorrectly. Compound words were translated incorrectly, because the dash character

mess up the model. The model cannot translate them separately and put them back together. The reason might be the training set does not contain a lot of compound word, so the model did not experience a lot with compound word. My made-up word failed the translation as well. This is because the model never see the same style/pattern of word before, so it has no experience at all. Also, "asuperlongwordistested" is very long, it is unlikely to be remember by the model. Thus, I expected the translation failed. I think the translation would be a little bit better if the model is trained more epochs. The successful translations show that the common/usual words are more likely to get translated correctly because the model see a lot of them and can easily do the translation. Please see the plot below for reference.

