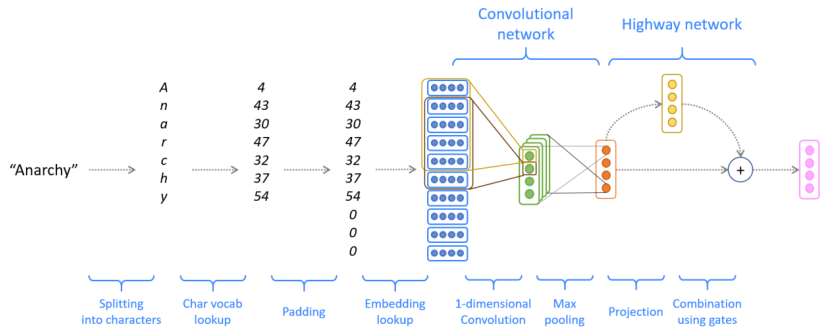
Assignment 5

Assignment 5
Part 1
(a)
(b)
(c)
(d)

Part 1



(a5.pdf)

(a)

> *why the embedding size used for character-level embeddings is typically lower than that used for word embeddings.*

We have to connect all the characters to a word. so if the size of e_{char} is big, the word vector will be huge and very expensive to compute.

(b)

> *Write down the total number of parameters in the character-based/word-based embedding model*

character-based

$num = e_{char} \times V_{char} + e_{word} \times e_{char} \times k + e_{word}$

word_based

$num = e_{word} \times V_{word}$

(c)

> *Explain one advantage of using a con-volutional architecture rather than a recurrent architecture for this purpose.*

It is much easier for CNN to compute in parallel, so it is much faster. Besides, CNN focuses more on part of the word, and RNN focus the whole word.

(d)

> *In lectures we learned about both max-pooling and average-pooling. For each*
> *pooling method, please explain one advantage in comparison to the other pooling method.*

max-pooling can discard something not so important and focus on the most important part.

average-pooling can know the whole information. (But sometimes it is bad to know all the stuff cause you cannot judge what you need to take out) (Less is more)

below (a5_updated.pdf)