

Image Caption Generation

201368087

April 25, 2020

Contents

1	Text Preparation	1
2	Network Outputs	1
3	Comparison of RNN vs LSTM	2
4	Additional Text Annotation Files	2

1 Text Preparation

One of the consequences of lemmatizing the words, means that any suffixes are removed from the words. This allows the model to focus on recognising objects within the image, and describing the object. Whereas without lemmatization, the model will have to also take into the account complex linguistical properties such as grammar, which is hard to model efficiently. Having to do two complex functions within one set of model and weights can be tricky, lemmatization can help to overcome that. Requiring the model to learn both grammar, and describing the image might go against abstraction principle; why should the model be required to do both tasks? It would make more sense for the model to just learn how to caption, and another program, possibly another model, could fix any grammatical mistakes. Especially since lemmatization is a very complex task, as it's not as simple as adding a suffix; in the case of words such as 'good' there is no 'gooder', as one example. Also, lemmatization keeps the vocabulary short, meaning that the model doesn't require to learn how each word can be used, and in which context. Mainly, the advantage of lemmatization is that it focusses the model's attention on relationship between words.

However, there are certain issues with lemmatization, as the generated captions will make often no grammatical sense, thus making it harder to evaluate, as there is no simple way to make the sentence be grammatically correct. However, leaving the sentence with grammatical errors, whilst achieving the goal of being close to the original captions, loses in terms of building coherent sentences a human would make. This makes it hard to say whether the model was successful or not. Also, at times lemmatization does not always lead to better performance, therefore it is best to not use it, unless the model's performance cannot be improved, thus trying lemmatization could help the performance. Furthermore, some sentiment analysis tools differentiate between different words based on the context and give a rating based on the context, thus lemmatization would be harmful if it were to be used in conjunction with tools like VADER.

To conclude, lemmatization could remove some useful words, or it could remove a lot of clutter thus leading to a better model. It is hard to say what impact lemmatization would have on the model, without trying it.

2 Network Outputs



Figure 1

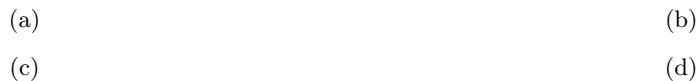


Figure 2

3 Comparison of RNN vs LSTM

4 Additional Text Annotation Files

There are two files; namely ExpertAnnotations.txt and CrowdFlowerAnnotations.txt. ExpertAnnotations.txt contains . In similar vein, the CrowdFlowerAnnotations.txt file contains

These files have similar functions, which solve a hard problem when it comes to models in this domain, which is encapsulating the wide variety of ways to describe a given image. Even if the model manages to successfully come close to the captions used in training, there are a variety of ways that describe the image just as well. Thus, the hardest part would be to get a large sample size to effectively narrow down all the possible ways of describing that object.

These files would help with this issue, as they could be used as a similarity ranking between the different forms of captions.

- an understanding of the contents of those two text files - an understanding of a key problem in this domain and how those additional files can help address the problem - your idea, justified, of how they could be incorporated into the model training and evaluation

References