in5400 Mandatory 2 Report

Bjørn-Andreas Lamo (janilam)
studentnummer: 575493

May 2021

1 Introduction

In this assignment we will train a model to caption photos. We will work with the COCO dataset[1]. Each photo has a series of true annotations of what the scene contains. There can be multiple different objects and object classes per image, but objects in the caption are easily recognizable. In addition the subset of the dataset we are working with is limited to top 10 feature detected, this is to save computational processing and time.

The predicted photo caption is compared with the true annotations through the BLEU-4 and METEOR linguistic score system.

Expected performance was achieved on all the models, except the on two layer LSTM attention; which reached only 25.55 METEOR score instead of the expected range 25.7 - 26.0

2 Report

2.1 Training parameters

Training parameters were unchanged from the base code for all models except for task 1 that had $num\_rnn\_layers = 1$. When $scheduler\_milestones$ and $scheduler\_factor$ is defined as below then the model will use an $MultiStepLR$ scheduler.

```
'optimizer': 'adamW',  # 'SGD' | 'adam' | 'RMSprop' | 'adamW'
'learningRate': {'lr': 0.001},  # learning rate to the optimizer
'weight_decay': 0.00001,  # weight_decay value
'number_of_cnn_features': 2048,  # Fixed, do not change
'embedding_size': 300,  # word embedding size
'vocabulary_size': 10000,  # number of different words
'truncated_backprop_length': 25,
'hidden_state_sizes': 512,  #
'num_rnn_layers': 2,  # number of stacked rnn's
'scheduler_milestones': [75,90], #45,70 end at 80? or 60, 80
'scheduler_factor': 0.2, #+0.25 dropout
```

2.2 Training

All models were trained with CUDA on UiO's HPC cluster.

Model	time/epoch
one layer RNN	00:45
two layer GRU	01:25
two layer LSTM	01:44
two layer att-LSTM	02:10

Below are graph over loss over epochs and the METEOR score.

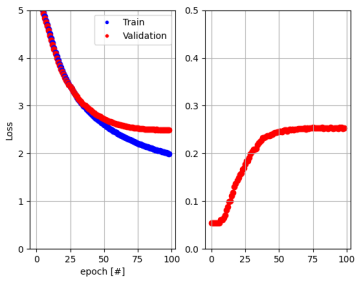

Figure 1: one layer RNN

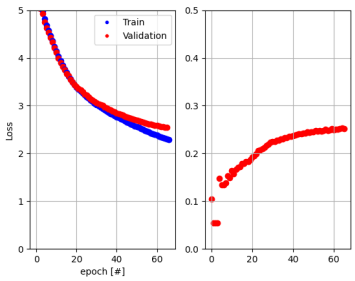

Figure 2: two layer GRU

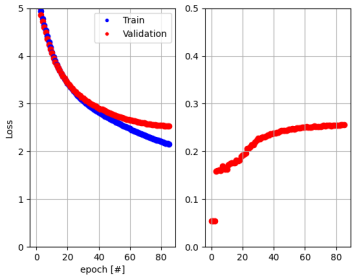

Figure 3: two layer LSTM

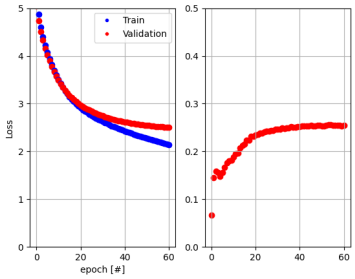

Figure 4: two layer LSTM with attention

2.3 Results

Top scores across epochs:

Model	BLEU-4	METEOR
one layer RNN	27.81	25.46
two layer GRU	27.93	25.51
two layer LSTM	**28.13**	**25.73**
two layer att-LSTM	27.45	25.55

The models meet expected scores, except for the two layer attention LSTM. This can be because the model didn't train as long as the other models, as the assignment text recommended fewer epochs.

2.4 Example predictions

Figure 5: a group of people that are standing in a field

Figure 6: a living room with a couch and a couch

Figure 7: a baseball player swinging a bat at a ball

Figure 8: a close up of a bowl of fruit in a bowl

Figure 9: a herd of cattle grazing on a lush green field

References

[1] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015.