

Assignment 3

W&B project link: [W&B Project Link](#)

Model architecture and chosen hyperparameters

random seed

RANDOM_SEED = 42

tokenizer params

TOKENIZER_NAME = "bert-base-uncased", MAX_SEQ_LENGTH = 256

datamodule params

VAL_SPLIT = 0.15, TRAIN_SPLIT = 0.7, TEST_SPLIT = 0.15, BATCH_SIZE = 32,
NUM_WORKERS = 8

model params

LSTM_LAYERS = 2, DROPOUT_RATE = 0.3, HIDDEN_DIM = 256,
EMBEDDING_DIM = 128, NUM_CLASSES = 2

training params

LEARNING_RATE = 1e-4, NUM_EPOCHS = 50, OPTIMIZER = "AdamW" # or some
other optimizer algorithm, SCHEDULER="ReduceLROnPlateau", LOSS="NLLoss"

Loss and accuracy curves from W&B



Final evaluation results

| Test metric | DataLoader 0 |
|-------------|---------------------|
| test_FN | 500.0 |
| test_FP | 656.0 |
| test_TN | 3057.0 |
| test_TP | 3287.0 |
| test_acc | 0.8458666801452637 |
| test_loss | 0.35022497177124023 |

Test Confusion Matrix:

| | 0 | 1 |
|---|------|------|
| 0 | 3057 | 656 |
| 1 | 500 | 3287 |

Example misclassifications and error analysis

3 misclassified examples:

Predicted: 1

True: 0

Text: spoiler warning!!! < br / > < br / > this is probably my least favourite cary grant film. i spent most of the time thinking : ignore him, leave him etc. is he trying to kill her? should we assume that after the end, when he persuades her to return home with him, that he will kill her after all? i see from checking in a couple of books that this is a reading that some viewers / critics have taken. finishing early is after all a more reasonable way to film a book, than to change the ending - and this is a more unreasonable change of ending than, for example, in altman's the long goodbye. unlike hitchcock's earlier example of making the murderer not the murderer because a star has been cast (when gosford park refers to the lodger, ivor novello's murderous habits are not mentioned) the book's ending is not definitely ruled out. < br / > < br / > in robin wood's hitchcock's films revisited, he writes : " as one mentally 'swiches' from ending to ending, the significance of every scene changes, the apparent solidity of the narrative dissolves, the illusion of the fiction's 'reality' "

Predicted: 0

True: 1

Text: i turned 13 when elvis hit the big times in 1956 with his first rca hit. a year later buddy holly stepped in to give the king some competition. one of buddy's major talents, besides his unique singing style and his songwriting ability, is often downplayed. buddy was also a skilled lead guitar player, developing a unique rockabilly style all his own on his fender strat. gary busey attempts to capture this aspect of buddy's persona. there were other contemporary master guitar rockers of equal caliber, such as chuck berry, carl perkins, and eddie cochrane, but buddy's talent is often overlooked. < br / > < br / > as noted by others, busey is the driving force behind the success of " the buddy holly story. " not only does the script play with the facts of buddy's life, but it even interjects several anachronisms for the two years of buddy's popularity, basically 1957 - 1959. one that comes to mind is the scene where buddy and maria are watching a 3 - d movie. buddy is disenchanted with it all and tells maria that it'll never last. it's just a fad. in reality there were no 3 - d movies

Predicted: 0

True: 1

Text: while browsing the internet for previous sale prices, i ran across these comments. why are they all so serious? it's just a movie and it's not pornographic. i acquired this short film from my parents 30 years ago and have always been totally delighted with it. i've shown it to many of my friends & they all loved it too. i feel privileged to own this original 1932 8mm black and white silent film of shirley before she became popular or well known. after reading the other comments, i agree that the film is " racy ". big deal! i only wish it was longer. it seems that i must be the only person who owns one of these originals, for sale at least, so i wonder how much it's worth?

Error Analysis:

The model struggles to discern the positivity or negativity of a review if it's not explicitly negative or positive. If words or possibly phrases indicative of positivity or negativity coincide in the same text it struggles to read the tone of the viewer.

Discussion and conclusion

Using a constant or a dynamic learning rate didn't make that much of a difference in our model's performance. We held all hyperparameters constant, except for the dropout rate and the hidden dimensions, to compare our model's behavior under a dynamic learning rate versus a constant learning rate. The experimental data indicated that our model made better predictions with a constant learning rate. When we changed to a dynamic learning rate and increased the dropout rate and hidden dimensions, the model performed somewhat worse than it did with a constant learning rate. We then, while still using a dynamic learning rate, tried holding constant the dropout rate and hidden dimensions. The model still underperformed relative to how it did with a constant learning rate. The model's performance did not improve dramatically and wasn't significantly different after switching from a constant learning rate to a dynamic learning rate. At the dynamic learning rate the model achieved a 0.85893 test accuracy; at a constant learning rate, 0.86973.

Citations

IMDB Dataset documentation

<https://docs.pytorch.org/text/stable/datasets.html#imdb>

Helps to understand the structure of the IMDB dataset

<https://www.kaggle.com/code/tusonggao/get-imdb-data-from-torchtext>

Python zip function

https://www.w3schools.com/python/ref_func_zip.asp

Pytorch LSTM

<https://docs.pytorch.org/docs/stable/generated/torch.nn.LSTM.html>

Pytorch Embedding

<https://docs.pytorch.org/docs/stable/generated/torch.nn.Embedding.html>

ReduceLROnPlateau

https://docs.pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html

torch.float(32)

<https://docs.pytorch.org/docs/stable/generated/torch.Tensor.float.html>